```
                    Copyright  GenCore version 5.1.6
                              (c) 1993 - 2005  Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on:          July 15, 2005, 10:36:59 ; Search time 8493 Seconds
                               (without alignments)
                               11022.663 Million cell updates/sec

Title:           US-10-736-868-1
Perfect score:   1932
Sequence:        1 atgattctattctatttt.........tgcctgcactgggattgtag 1932

Scoring table:   IDENTITY_NUC
                 Gapop 10.0 , Gapext 1.0

Searched:        4708233 seqs, 2427607955 residues

Total number of hits satisfying chosen parameters:      9416466

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       GenEmbl:*
                  1: gb_ba:*
                  2: gb_htg:*
                  3: gb_in:*
                  4: gb_om:*
                  5: gb_ov:*
                  6: gb_pat:*
                  7: gb_ph:*
                  8: gb_pl:*
                  9: gb_pr:*
                 10: gb_ro:*
                 11: gb_sts:*
                 12: gb_sy:*
                 13: gb_un:*
                 14: gb_vi:*
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	Query Match %	Length	DB	ID	Description
	1	1932	100.0	2131	3	AY360470	AY360470 Caenorhab
	2	894	46.3	45389	3	U80032	U80032 Caenorhabdi
c	3	426.4	22.1	17418	3	CBRG03O12	CBRG03O12 Caenorhab
c	4	211	10.9	39089	3	CBRG33D04	CBRG33D04 Caenorhab
c	5	75.4	3.9	7218	6	I66494	I66494 Sequence 14
c	6	57.2	3.0	2000	9	AX655393	AX655393 Sequence
	7	56.2	2.9	161964	2	AC111055	AC111055 Mus muscu
c	8	56	2.9	236428	10	AC147563	AC147563 Mus muscu
	9	55.4	2.9	171583	10	AC116690	AC116690 Mus muscu
	10	54.8	2.8	346939	1	BX842643	BX842643 Mycoplasm
c	11	54.6	2.8	196150	10	AC147250	AC147250 Mus muscu
c	12	54.6	2.8	252384	10	AC123071	AC123071 Mus muscu
	13	54	2.8	220873	2	AC120258	AC120258 Rattus no
c	14	54	2.8	225493	2	AC097214	AC097214 Rattus no
	15	54	2.8	225983	2	AC113937	AC113937 Mus muscu
c	16	54	2.8	237283	2	AC111786	AC111786 Rattus no
	17	53.6	2.8	1083	3	AY044180	AY044180 Plasmodiu
	18	53.6	2.8	168958	10	AL591712	AL591712 Mouse DNA
c	19	53.6	2.8	229052	2	AC116514	AC116514 Mus muscu
c	20	53.6	2.8	239604	2	AC111294	AC111294 Rattus no
c	21	53.6	2.8	252644	2	AC097392	AC097392 Rattus no
	22	53.6	2.8	262724	2	AC114430	AC114430 Rattus no
	23	53.2	2.8	179772	2	AC141474	AC141474 Mus muscu
c	24	53	2.7	175030	10	AC129942	AC129942 Mus muscu
c	25	53	2.7	185912	2	AC141221	AC141221 Rattus no
c	26	53	2.7	203877	10	AC122260	AC122260 Mus muscu
c	27	53	2.7	228137	10	AL845323	AL845323 Mouse DNA
	28	52.8	2.7	62474	2	AC137973	AC137973 Mus muscu
c	29	52.8	2.7	185849	5	BX005073	BX005073 Zebrafish
c	30	52.4	2.7	180642	10	AC132335	AC132335 Mus muscu
	31	52.4	2.7	198414	2	BX004767	BX004767 Mus muscu
	32	52.2	2.7	158504	2	AC107810	AC107810 Mus muscu
	33	52.2	2.7	161173	2	AC110881	AC110881 Mus muscu
	34	52.2	2.7	199554	10	AC123883	AC123883 Mus muscu
c	35	52.2	2.7	201219	10	AC145198	AC145198 Mus muscu
c	36	52.2	2.7	208505	10	AC122247	AC122247 Mus muscu
c	37	52	2.7	188817	10	AC131297	AC131297 Mus muscu
c	38	52	2.7	189048	10	AL672221	AL672221 Mouse DNA
	39	52	2.7	264653	2	AC113705	AC113705 Rattus no
c	40	51.8	2.7	179371	10	AC140335	AC140335 Mus muscu
	41	51.6	2.7	1950	1	AY142081	AY142081 Borrelia
	42	51.6	2.7	2196	1	BBU78764	U78764 Borrelia bu
	43	51.6	2.7	2261	1	BBU72996	U72996 Borrelia bu
	44	51.6	2.7	4283	1	BBU44912	U44912 Borrelia bu
	45	51.6	2.7	30750	1	AE001575	AE001575 Borrelia

ALIGNMENTS

```
RESULT 1
AY360470
LOCUS       AY360470                2131 bp    mRNA    linear   INV 01-JUN-2004
DEFINITION  Caenorhabditis elegans osmotic stress resistance protein (osr-1)
            mRNA, osr-1-rm1 allele, complete cds.
ACCESSION   AY360470
VERSION     AY360470.1  GI:38154581
KEYWORDS    .
SOURCE      Caenorhabditis elegans
  ORGANISM  Caenorhabditis elegans
            Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
            Rhabditoida; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE   1  (bases 1 to 2131)
  AUTHORS   Solomon,A., Bandhakavi,S., Jabbar,S., Shah,R., Beitel,G.J. and
            Morimoto,R.I.
  TITLE     Caenorhabditis elegans OSR-1 Regulates Behavioral and Physiological
            Responses to Hyperosmotic Environments
  JOURNAL   Genetics 167 (1), 161-170 (2004)
   PUBMED   15166144
REFERENCE   2  (bases 1 to 2131)
  AUTHORS   Solomon,A., Bandhakavi,S., Jabbar,S., Shah,R., Beitel,G.J. and
            Morimoto,R.I.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-AUG-2003) Biochemistry, Northwestern University, 2205
            Tech Drive, Hogan 2-100, Evanston, IL 60201, USA
FEATURES             Location/Qualifiers
     source          1..2131
                     /organism="Caenorhabditis elegans"
                     /mol_type="mRNA"
                     /db_xref="taxon:6239"
                     /chromosome="I"
                     /map="-0.2 m.u."
     gene            1..2131
                     /gene="osr-1"
                     /allele="osr-1-rm1"
     5'UTR           1..98
                     /gene="osr-1"
                     /allele="osr-1-rm1"
     CDS             99..2030
                     /gene="osr-1"
                     /allele="osr-1-rm1"
                     /function="mediates multiple responses to osmotic stress"
```

us-10-736-868-1.rge

```
           /note="OSR-1"
           /codon_start=1
           /product="osmotic stress resistance protein"
           /protein_id="AAR12229.1"
           /db_xref="GI:3815482"
           /translation="MILFLFLFLLLGFCIAPLSAQSPSTSDAPGALLSLVGKSHQKL
           PLAPSMEALELMGVQFVDALIKKGQMEMAKGAFKTQLEVLEKVHPDQFPKYKKLKVDD
           LAADAVMQQAEMAKLQPKSGNAFIDMLNGNGIPIGSSIRGLEDAIRTQRDMENTDPSE
           QIAKAVMDKFQTQLPLGLVANMIAGKNPFKMPQQMRKAQAAPSSVFQQALAQRAMLGK
           NAPVAGGREEQRMMNRVDQRMQQRELQEEDEDDDLEDEDVPRRRSSDGPQSEAE
           HQRRDLARRLKSSPRLKELLQNAEVQSLLSYQRMRDSPLSKRRPLAMNDEDESAFRAM
           EARAKLDQKSQLVIGLHGFGESDDDEDEDENLIDPSENSFRRAPLRLSSGFVEKLKS
           NDELKSALDRIKYRVDDVEKYLAPKMEFNPKPOGYFAPRKIPTPRRKMLPLLIGSD
           PKVQEEIRRHPSTEWKIAKESRVLTNLKNNPSLAALFMDDKLENTLKGRQMLTDEQKG
           RTRVKTIRALPRLFGAPTAKAEMIDAKVFQDIEERPIPPLFFEPKGRHTLRRWTGANE
           KEIPGLGSRFIPLSLDPTMPALNTAFSTQGRARDEWDTMFKIPNNWNPGDEVGFKMNS
           KTKRFPVGNGAFDMPALGL"
3'UTR      2031..2131
           /gene="osr-1"
           /allele="osr-1-rm1"

ORIGIN

Query Match         100.0%;  Score 1932;  DB 3;  Length 2131;
Best Local Similarity   100.0%;
Matches 1932;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy      1 ATGATTTTATTTTATTTTTATTCCTGTTGTTAGGATTTTTGTATCGCACCATTATCGGCC 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     99 ATGATTTTATTTTATTTTTATTCCTGTTGTTAGGATTTTTGTATCGCACCATTATCGGCC 158

Qy     61 CAATCTCCATCCGACTTCCGATGCTCCGGGAGCTTGTTGTTGTCATCTCTCGTAGGTAAAAGC 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    159 CAATCTCCATCCGACTTCCGATGCTCCGGGAGCTTGTTGTTGTCATCTCTCGTAGGTAAAAGC 218

Qy    121 CATCAAAAACTACCACTGGCTCCATCAATGGAAGCTCTTGAACTGATGGTGTTCAATTT 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    219 CATCAAAAACTACCACTGGCTCCATCAATGGAAGCTCTTGAACTGATGGGTGTTCAATTT 278

Qy    181 GTTGATGCTCTCATCAAAAAGGTCAAAATGCAAAATGGAGGAGCATTAAGACTCAA 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    279 GTTGATGCTCTCATCAAAAAAGTCAAAATGCAAAATGGAGGAGCATTAAGACTCAA 338

Qy    241 TTAGAAGTTCTAGAGAAAGTACATCCTGATCAATTCGATAAGTACAAAAGCTAAAAGTT 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    339 TTAGAAGTTCTAGAGAAAGTACATCCTGATCAATTCGATAAGTACAAAAGCTAAAAGTT 398

Qy    301 GATGATTTGGCAGCTGAGCTGATGCAACAGGCGGAGATGGCAAAATTACAGCCTAAA 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    399 GATGATTTGGCAGCTGAGCTGATGCAACAGGCGGAGATGGCAAAATTACAGCCTAAA 458

Qy    361 TCAGGAAATGCATTTATCGATATGTTGAATGGAATCCCAATTGGTAGCAGTATT 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    459 TCAGGAAATGCATTTATCGATATGTTGAATGGAATCCCAATTGGTAGCAGTATT 518

Qy    421 CGTGGATTAGAAGATGCTATCCGAACGCAGAGACATGGAAAATACGGATCCGTCCGAA 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    519 CGTGGATTAGAAGATGCTATCCGAACGCAGAGACATGGAAAATACGGATCCGTCCGAA 578

Qy    481 CAGATTGCCAAAGCCGTAATGGACAAAATTTCAAACACAATTCTCCGAAAAATGCC 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    579 CAGATTGCCAAAGCCGTAATGGACAAAATTTCAAACACAATTCTCCGAAAAATGCC 638

Qy    541 AATATGATCGCTGGCAAGAACCCCTTTAAAAATGCCTCAACAAATGAGAAAAGCTCAAGCT 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    639 AATATGATCGCTGGCAAGAACCCCTTTAAAAATGCCTCAACAAATGAGAAAAGCTCAAGCT 698

Qy    601 GCTCCATCGTCAGTGTTCCAACAAGCTCTTGCTCAAAGAGCAATGTTAGGTAAAAATGCC 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    699 GCTCCATCGTCAGTGTTCCAACAAGCTCTTGCTCAAAGAGCAATGTTAGGTAAAAATGCC 758

Qy    661 CCAGTTGCCGGTGGAAGAGGTGAAGAACAACGGATGATGAATCGAGTCGACCAAAGA 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    759 CCAGTTGCCGGTGGAAGAGGTGAAGAACAACGGATGATGAATCGAGTCGACCAAAGA 818

Qy    721 ATGCAACAAAGAGAACTTCAAGAGGAAGAAGATGAAGAGATGATGAGATGATCTTGAGGACGAGGAT 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    819 ATGCAACAAAGAGAACTTCAAGAGGAAGAAGATGAAGAGATGATGAGATGATCTTGAGGACGAGGAT 878

Qy    781 GTACCCAGAGAAGAGAGTTCGGATGGAGAACCACCAAAGTGAAGCAGCATCAGAGAAGA 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    879 GTACCCAGAGAAGAGAGTTCGGATGGAGAACCACCAAAGTGAAGCAGCATCAGAGAAGA 938

Qy    841 GATTTAGCCAGGAGATTGAAAAGTAGTCCTAGATTAAAAGAGCTTTTACAGAATGCGAA 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    939 GATTTAGCCAGGAGATTGAAAAGTAGTCCTAGATTAAAAGAGCTTTTACAGAATGCGAA 998

Qy    901 GTTCAATCATTGCTCTCTTACCAACCAATGAGGGATTCTCCACTGAGCAAGCGAAGGCCT 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    999 GTTCAATCATTGCTCTCTTACCAACCAATGAGGGATTCTCCACTGAGCAAGCGAAGGCCT 1058

Qy    961 TTGGCTATGAACGATGAGGATGAAAGTGCATTCCGCGCAATGGAGGTCGTGCAAAACTA 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1059 TTGGCTATGAACGATGAGGATGAAAGTGCATTCCGCGCAATGGAGGTCGTGCAAAACTA 1118

Qy   1021 GATCAAAAATCTCAACTTGTCGCTCCATGGTTTTTGGAGAGTCTGATGATGATGATGAA 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1119 GATCAAAAATCTCAACTTGTCGCTCCATGGTTTTTGGAGAGTCTGATGATGATGATGAA 1178

Qy   1081 GACGAAGAAGATGAAAATTTGATTGATCATCTGAAAATTCATTCCGTCGTGCGTCCACCTT 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1179 GACGAAGAAGATGAAAATTTGATTGATCATCTGAAAATTCATTCCGTCGTGCGTCCACCTT 1238

Qy   1141 CGTCTTCTTCCGGATTCGTTGGTGAGAAATAAAGTCAAATGAATGATGAAAAGTGCATTG 1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1239 CGTCTTCTTCCGGATTCGTTGGTGAGAAATAAAGTCAAATGAATGATGAAAAGTGCATTG 1298

Qy   1201 GACAGAATTAAATATCGAGTTGATGACGTGAAAAGTATCTTGCTCCAAAGCCGATGGAA 1260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1299 GACAGAATTAAATATCGAGTTGATGACGTGAAAAGTATCTTGCTCCAAAGCCGATGGAA 1358

Qy   1261 TTCAATCCAAAACCTCAGCCTGGCTACTTTGCTCCACGTAAAATCCCAACAAGACCACCGT 1320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1359 TTCAATCCAAAACCTCAGCCTGGCTACTTTGCTCCACGTAAAATCCCAACAAGACCACCGT 1418

Qy   1321 AAAAATGCTTCATTATTAATTGGATCATCTGATCAAGAGAAGTTCAAGAGAAATGAAGACAT 1380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1419 AAAAATGCTTCATTATTAATTGGATCATCTGATCAAGAGAAGTTCAAGAGAAATGAAGACAT 1478

Qy   1381 CCAAGTACCGAATGGAAAATTGCAAAAGAATCAAGAGTTTTGACAAATTTGAAGAATAAT 1440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1479 CCAAGTACCGAATGGAAAATTGCAAAAGAATCAAGAGTTTTGACAAATTTGAAGAATAAT 1538

Qy   1441 CCAAGTCTTGCTGCATTGTTCATGGATGATGATAAAATTAGAGAATACAATTGAAAGGAAGGCAA 1500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1539 CCAAGTCTTGCTGCATTGTTCATGGATGATGATAAAATTAGAGAATACAATTGAAAGGAAGGCAA 1598

Qy   1501 ATGTTAACTGATGAACAGAGAAGGTAGAACAACGTGTCAAAACAATTCGTGCATTACCAAGA 1560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1599 ATGTTAACTGATGAACAGAGAAGGTAGAACAACGTGTCAAAACAATTCGTGCATTACCAAGA 1658

Qy   1561 CTGTTCGGTGACCAACTGCAAAAGTGAAATGATTGAGAAGTATTCAAGATATT 1620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1659 CTGTTCGGTGACCAACTGCAAAAGTGAAATGATTGAGAAGTATTCAAGATATT 1718

Qy   1621 GAAGAACGTCCCATTCCTCCATTGTTCTTGAACCAAAAGAAGGCATACAGAGATTGAGA 1680
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1719 GAAGAACGTCCCATTCCTCCATTGTTCTTGAACCAAAAGAAGGCATACAGAGATTGAGA 1778

Qy   1681 TGGACTGGAGCAAATGAAATGCCTCAAGAGCCAATGTTAGGTAAAAATGCC 1740
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1779 TGGACTGGAGCAAATGAAATGCCTCAAGAGCCAATGTTAGGTAAAAATGCC 1838

Qy   1741 CTTGATCAACTATGCCAGCCTTGAACACGGCTTTCTGACTCAGGGCCGAGCCCGTGAC 1800
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1839 CTTGATCAACTATGCCAGCCTTGAACACGGCTTTCTGACTCAGGGCCGAGCCCGTGAC 1898
```

```
Qy   1801 GAGTGGGATACCATGTTCAAAATCCCGAATAACTGGAATCCTGGAGATGAAGTTGGGTTC 1860
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1899 GAGTGGGATACCATGTTCAAAATCCCGAATAACTGGAATCCTGGAGATGAAGTTGGGTTC 1958

Qy   1861 AAAATGAACTCAAAAACCAAACGATTCGTTGGAGGAAATGGAGCATTTGATATGCCTGCA 1920
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1959 AAAATGAACTCAAAAACCAAACGATTCGTTGGAGGAAATGGAGCATTTGATATGCCTGCA 2018

Qy   1921 CTGGGATTGTAG 1932
          ||||||||||||
Db   2019 CTGGGATTGTAG 2030
```

RESULT 2
U80032/c
LOCUS   U80032   45389 bp   DNA   linear   INV 16-SEP-2004
DEFINITION Caenorhabditis elegans cosmid C32E12, complete sequence.
ACCESSION   U80032
VERSION   U80032.1 GI:2073526
KEYWORDS   HTG.
SOURCE   Caenorhabditis elegans
  ORGANISM Caenorhabditis elegans
            Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
            Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE   1 (bases 1 to 45389)
  CONSRTM   WormBase Consortium
  TITLE   Genome sequence of the nematode C. elegans: a platform for
            investigating biology. The C. elegans Sequencing Consortium
  JOURNAL   Science 282 (5396), 2012-2018 (1998)
   MEDLINE   99069613
   PUBMED   9851916
REFERENCE   2 (bases 1 to 45389)
  AUTHORS   Wilcox,L.
  TITLE   The sequence of C. elegans cosmid C32E12
  JOURNAL   Unpublished (2001)
REFERENCE   3 (bases 1 to 45389)
  AUTHORS   Waterston,R.
  TITLE   Direct Submission
  JOURNAL   Submitted (27-NOV-1996) Department of Genetics, Washington
            University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
            Louis, MO 63110, USA
REFERENCE   4 (bases 1 to 45389)
  AUTHORS   Waterston,R.
  TITLE   Direct Submission
  JOURNAL   Submitted (08-MAY-1997) Department of Genetics, Washington
            University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
            Louis, MO 63110, USA
REFERENCE   5 (bases 1 to 45389)
  AUTHORS   Waterston,R.
  TITLE   Direct Submission
  JOURNAL   Submitted (18-OCT-2001) Department of Genetics, Washington
            University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
            Louis, MO 63110, USA
REFERENCE   6 (bases 1 to 45389)
  AUTHORS   Waterston,R.
  TITLE   Direct Submission
  JOURNAL   Submitted (19-APR-2002) Department of Genetics, Washington
            University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
            Louis, MO 63110, USA
REFERENCE   7 (bases 1 to 45389)
  AUTHORS   Waterston,R.
  TITLE   Direct Submission
  JOURNAL   Submitted (20-JUL-2002) Department of Genetics, Washington
            University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
            Louis, MO 63110, USA
REFERENCE   8 (bases 1 to 45389)
  AUTHORS   Waterston,R.
  TITLE   Direct Submission
  JOURNAL   Submitted (19-NOV-2002) Department of Genetics, Washington
            University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
            Louis, MO 63110, USA
REFERENCE   9 (bases 1 to 45389)
```

AUTHORS   Waterston,R.
TITLE     Direct Submission
JOURNAL   Submitted (13-JAN-2003) Department of Genetics, Washington
          University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
          Louis, MO 63110, USA
REFERENCE   10 (bases 1 to 45389)
  AUTHORS   WormBase Consortium
  TITLE     Direct Submission
  JOURNAL   Submitted (16-SEP-2004) Department of Genetics, Washington
            University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
            Louis, MO 63110, USA
COMMENT     On May 8, 1997 this sequence version replaced gi:1703546.
            Submitted by:
              Genome Sequencing Center
              Department of Genetics, Washington University
              St. Louis, MO 63110, USA, and
              Sanger Centre, Hinxton Hall
              Cambridge CB10 1RQ, England
              email: submissions@watson.wustl.edu and jes@sanger.ac.uk

            NOTICE: This sequence may not be the entire insert of this clone.
            It may be shorter because we only sequence overlapping sections
            once, or longer because we provide a small overlap between
            neighboring submissions.

            This sequence was finished as follows unless otherwise noted: all
            regions were double stranded, sequenced with an alternate chemistry
            or covered by high quality data (i.e., phred quality >= 30); an
            attempt was made to resolve all sequencing problems, such as
            compressions and repeats; all regions were covered by sequence from
            more than one m13 subclone.

            For a graphical representation of this clone sequence and its
            analysis see:
            http://www.wormbase.org/db/seq/sequence?name=C32E12;class=Sequence

                         NEIGHBORING CLONE INFORMATION

            The 5' clone is F56A3, 200 bp overlap; the 3' clone is B0261, 3201
            bp overlap. Actual start of this clone is at base position 1 of
            C32E12; actual end is at 42388 of C32E12.

            NOTES:

            Coding seqences below are the result of integration and manual
            review of the following data : computer analysis using the program
            Genefinder (P. Green and L. Hillier, personal communication), the
            large scale EST projects of Yuji Kohara
            (http://www.ddbj.nig.ac.jp/c-elegans/html/CE_INDEX.html) and The C.
            elegans ORFeome cloning project (http://worfdb.dfci.harvard.edu/),
            similarity to other proteins from BlastX analyses
            (http://blast.wustl.edu/), sequence conservation with C. briggsae
            using Jim Kent's WABA alignment program (Genome Research
            10:1115-1125, 2000), individual C. elegans GenBank submissions,
            and personal communications with C. elegans researchers. tRNAs
            are predicted using the program tRNAscan-SE (Lowe, T.M. and
            Eddy, S.R., 1997, Nucl. Acids. Res., 25, 955-964).
FEATURES             Location/Qualifiers
     source          1..45389
                     /organism="Caenorhabditis elegans"
                     /mol_type="genomic DNA"
                     /strain="Bristol N2"
                     /db_xref="taxon:6239"
                     /chromosome="I"
                     /clone="C32E12"
     misc_RNA        750..772
                     /gene="mir-88"
                     /product="micro RNA"
                     /note="standard_name=C32E12.6"
                     /db_xref="WormBase:C32E12.6"
     gene            972..2572
                     /gene="C32E12.2"

```
Qy    29 TGTTAGGATTTGTATCGCACCATTATCGGCCCAATCTCCATCGACTTCCGATGCTCCGG   88
            ||||  ||| ||||||||||||||||||||||||||||||||||||||||||||
Db  6077 TATTTAGAAAATTTCCAGCACCATTATCGGCCCAATCTCCATCGACTTCCGATGCTCCGG 6018

Qy    89 GAGCTTTGTTGTCATCTCTCGTAGGTAAAAGCCATCAAAAACTACCACTGGCTCCATCAA  148
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  6017 GAGCTTTGTTGTCATCTCTCGTAGGTAAAAGCCATCAAAAACTACCACTGGCTCCATCAA 5958

Qy   149 TGGAAGCTCTTGAACTGATGGGTGTTCAATTTGTTGATGCTCTCATCAAAAAAGGTCAAA  208
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  5957 TGGAAGCTCTTGAACTGATGGGTGTTCAATTTGTTGATGCTCTCATCAAAAAAGGTCAAA 5898

Qy   209 TGGAAATGGCAAAAGGAGACATTTAGACTCAATTGAAGTTCTAGAGAAGAAACTACATCCTG 268
         |||||||||||||||||||||||||||||||||||||||||||||||  |||||||||||
Db  5897 TGGAAATGGCAAAAGGAGACATTTAGACTCAATTGAAGTTCTAGAGAAAGACTACATCCTG 5838

Qy   269 ATCAATTCGATAAGTACAACAAAAGCTAAAAGTTGATGATTGGCAGCTGATGCAGTTATGC 328
         |||||||||||||||||||  |||||||||||||||||||||||||||||||||||||||
Db  5837 ATCAATTCGATAAGTACAAAAAGCTAAAAGTTGATGATTGGCAGCTGATGCAGTTATGC 5778

Qy   329 AACAGGCGGAGATGGCAAAATTACAGCCTAAA 360
         ||||||||||||||||||||||||||||||||
Db  5777 AACAGGCGGAGATGGCAAAATTACAGCCTAAATCAGGTTTTCATTTTACTTTTTCAGTAA 5718

Qy   361 ------------------------------TCAGGTTTTCATTTTACTTTTTCAGTGAAAT 396
                                       |||||||||||||||||||||||||
Db  5717 CTGTTCTGAAAACAAATTTGAATTTCAGGTTTTCATTTTACTTTTTCAGTGAAAT 5658

Qy   397 GGAATCCCAATTGGTAGCAGTATTCGTGGATTAGAAGCAATGATGTCATCCCAACCAGAGAGAT 456
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  5657 GGAATCCCAATTGGTAGCAGTATTCGTGGATTAGAAGCAATGATGTCATCCCAACCAGAGAGAT 5598

Qy   457 ATGGAAAATACGGATCCGTCCGAAC 481
         |||||||||||||||||||||||||
Db  5597 ATGGAAAATACGGATCCGTCCGAACAGCAGGTTGCGTGTGAATTCAATTGCCAGTTTTAAAAAT 5538

Qy   482 ------------------------------------------------------------ 481
Db  5537 ATTAGAGAAACTCCCCTTTCCCATTGTTGTTTCCCATTGTTTTTAATTATTGTTGCAGAATGATT 5478

Qy   482 ------------------------------AGATTGCCAAAGCCGTAATGGACAAA 507
                                       ||||||||||||||||||||||||||
Db  5477 TTCTTAATTTTATAACTAACACATGTTTCTCTAATTTAGATTGCCAAAGCCGTAATGGACAAA 5418

Qy   508 TTTCAAACACAATTCTCCAGGACTCGTTGCAAATGATCGCTGGCAAGAACCCCTTT 567
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  5417 TTTCAAACACAATTCTCCAGGACTCGTTGCAAATGATCGCTGGCAAGAACCCCTTT 5358

Qy   568 AAAAATGCCTCAACAAATGAGAAAAGCTCAAGCTGTTCCATCGTGTTCCAACAAGCT 627
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  5357 AAAAATGCCTCAACAAATGAGAAAAGCTCAAGCTGTTCCATCGTGTTCCAACAAGCT 5298

Qy   628 CTTGCTCAAAGAGAGCAATGTTAGGTAGGAAGAAAAATGCCCCAGTTGCCGGTGGTGAAGAGGTGAAGAA 687
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  5297 CTTGCTCAAAGAGAGCAATGTTAGGTAGGAAGAAAAATGCCCCAGTTGCCGGTGGTGAAGAGGTGAAGAA 5238

Qy   688 CAACGGATGATGATGAATCGAGTGGACCAACAAAGAATGCAACAAAGAGACTTCAAGAGGAA 747
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  5237 CAACGGATGATGATGAATCGAGTGGACCAACAAAGAATGCAACAAAGAGACTTCAAGAGGAA 5178

Qy   748 GATGAAGATGATGATGATCTTGAGGACGACCCAGTACCCAGAGAAGTTCGGATGGA 807
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  5177 GATGAAGATGATGATGATCTTGAGGACGACCCAGTACCCAGAGAAGTTCGGATGGA 5118

Qy   808 GAACCACAAAGTGAACGAGCAGCATCAGAGAGAAGATTTAGCCAGGAGATTGAAAAGTAGT 867
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  5117 GAACCACAAAGTGAACGAGCAGCATCAGAGAGAAGATTTAGCCAGGAGATTGAAAAGTAGT 5058

Qy   868 CCTAGATTAAAAGAGCTTTA 888
         ||||||||||||||||||||
Db  5057 CCTAGATTAAAAGAGCTTTTACAGGTGAATTAGTAATTTCTCAACTGTTTTAAATTTTTT 4998
```

CDS
  /locus_tag="C32E12.2"
  join(972..1126,1184..1438,1501..1587,1633..1786,1997..2325,2371..2572)
  /gene="C32E12.2"
  /standard_name="C32E12.2"
  /note="contains similarity to Pfam domains PF01590 (GAF domain), PF00233 (3'5'-cyclic nucleotide phosphodiesterase); coded for by the following C. elegans cDNAs: OSTF082F3_1, OSTR082F3_1"
  /codon_start=1
  /product="Hypothetical protein C32E12.2"
  /protein_id="AAB53879.1"
  /db_xref="GI:1703547"
  /db_xref="WormBase:C32E12.2"
  /translation="MSKGIAGYVASTGEGLNIENAYEDERFNADVDSKTGYTKTILC
  MPILIRGIVIGVVQMVKHDGVFTRQDEDAFEIFAVYCGLALHHAKLYDKIRRSEQKY
  RVALEVLAYHSVCNADEVNKLKKIEINNRIVELETIDFNGMRLSELEKPLYAVYMFKT
  LFADTLRFDTEDLIRFVLTVRKNYRVRVAYHNWAHGWSVAHAMFATLMNSPDAFTKLEA
  LALYVSCLCHDLDHRGKNNAYMKTMSTPLASIYSTSVMERHHFNQTVTILQQDGHNIL
  KSLSSEDYKKTLSLIKHCILATDLALFFSNKAKLNVILDNNTFDINRQEHRLLTQAVM
  MTGCDLVASAKPWNIQTETVKVIPEBFYDQGDAERLSGKEPIPMDRQQAHMLPQMQA
  S"

gene
  complement(3240..6294)
  /gene="osr-1"

CDS
  complement(join(3240..3454,3601..3712,3759..4390,4446..4527,5034..5441,5571..5689,5742..6059,6249..6294))
  /locus_tag="C32E12.3"
  /gene="osr-1"
  /standard_name="C32E12.3"
  /note="contains similarity to Schizosaccharomyces pombe Hypothetical protein C29B6.10c in chromosome I.; SW:YAFA_SCHPO; coded for by the following C. elegans cDNAs: yk563c9.3, AY360470"
  /codon_start=1
  /product="Osmotic stress resistant protein 1"
  /protein_id="AAM15555.1"
  /db_xref="GI:20198782"
  /db_xref="WormBase:C32E12.3"
  /translation="MILFLFLLLGFCIAPLSAQSPSTSDAPGALLSSLVGKSHQKL
  PLAPSMEALELMGVQVFVDALIKKGQMEMAKGAFKTQLEVLEKVHPDGFKYKKLKVDD
  LAADAVMQQAEMAKLQPKSGNAFIDMLNGNGIPIGSSIRGLEDAIRTQRDMENTDPSE
  QIAKAVMDKFQTQILPGLVANMIAGKNPFKMPQQMRKAQAAPSSVFQQALAQRAMLGK
  NAPVAGGRGEEQRMMNRVDQRMQQRELQEEDEDDDLEDEDVPRRRSSDGEFQSEAE
  HQRRDLAARLKSSPRLKELLQNAFVQSLLSYQRMRDSPLSKRRPLAMNDEDESAFRAM
  EARAKLDQKSQLVLGLHGFGESDDDEDEDENLIDPSENSFRRAPLRLSSGFVEKLKS
  NDELKSALDRIKYRVDDVEKYLAPKPMEFNPKFQPGYFAPRKIPTRPRKMLPLLIGSD
  PKVQEIRRHPSTEWKIAXESRVLTNLKNNPSLAALFMDDKLENTLKGRQMLTDEQKG
  RTRVKTIRALPRLFGAPTAKAEMIDAKVFQDIEERPIPPLFFEPKGRHTRLRWTGANE
  KEIPGLGSRFILPSLDPTMPALNTAFSTQGRARDEWDTMFKIPNNWNPGDEVGFKMNS
  KTKRFVGGNGAFDMPALGL"

gene
  complement(8197..11009)
  /gene="sox-1"

CDS
  complement(join(8197..8424,9119..9232,9281..9688,10208..10471,10809..11009))
  /locus_tag="C32E12.5"
  /gene="sox-1"
  /standard_name="C32E12.5"
  /note="contains similarity to Pfam domain PF00505 (HMG (high mobility group) box); coded for by the following C. elegans cDNAs: yk397g12.5, yk404e6.5, yk657g12.5, yk1577b07.5, yk1661e08.5, yk397g12.3, yk404e6.3, yk657g12.3, yk1577b07.3, yk1661e08.3"
  /codon_start=1
  /product="Sox (mammalian sry box) family protein 1"
  /protein_id="AAM15556.1"
  /db_xref="GI:20198784"

Query Match   46.3%; Score 894; DB 3; Length 45389;
Best Local Similarity   66.7%; Pred. No. 1.9e-195;
Matches 1894; Conservative   0; Mismatches 10; Indels 934; Gaps   6;

```
Qy    889 ----------------------------------------------------------- 888
Db   4997 TAGAAAGGAAAAAAATCCAAATATTTAGGCTTTATATAAAAGTATTAAGTCTTTTTAT 4938

Qy    889 --------------------------------------------------------- 888
Db   4937 ATGTATTTAGAGCATATCTGTAAATATATAATTTACAAATACAACAATCATACATAGTTGT 4878

Qy    889 --------------------------------------------------------- 888
Db   4877 TGAAGAATCGCCGATATTGATAACTTCAACAGTCGTAATACTTCTGCAGCAATTATCAA 4818

Qy    889 --------------------------------------------------------- 888
Db   4817 AAAAATTGCCGAATCTTACTATTTTAAGTCTATATTTTTCGGAAACTTCTAAAAAAGT 4758

Qy    889 --------------------------------------------------------- 888
Db   4757 TTTTTGTCATTTTGAAAATTTAATATATAATAGGATGCGATTTAGCTAAAGTCTCCTATTTT 4698

Qy    889 --------------------------------------------------------- 888
Db   4697 GTTTGTAAGATGTTAAACCACCATAAAAATATTTCGAAAAAATATTTGTGGAGATCGCGAAT 4638

Qy    889 --------------------------------------------------------- 888
Db   4637 TGAGAAAAAATCACGTGTATCTTGGAAAATAACTAACAATGACAGTCGTAAGCCTAGTAG 4578

Qy    889 ------------------------------------CAGAATGCGGAAG 901
Db   4577 GACATTTGTGCCTGTGGTCGCTTGAAAAAATACCTAATACTTTCAGAATGCGGAAG 4518

Qy    902 TTCAATCATTGCTCTCTTACCACGAATGAGGATTCTCCACTGAGCAAGCGGAAGGCCTT 961
Db   4517 TTCAATCATTGCTCTCTTACCACGAATGAGGATTCTCCACTGAGCAAGCGGAAGGCCTT 4458

Qy    962 TGGCTATGAAC--------------------------------------------- 972
Db   4457 TGGCTATGAACGGTTGGTTTTTGTTTTTGAATTAGTAGGAACTCTTATCTGTCGACTAA 4398

Qy    973 ------GATGAGGATGAAAGTGCATTCGGCGCAATGCAAATCTAGATCAA 1026
Db   4397 ATTTCAGATGAGGATGAAAGTGCATTCGGCGCAATGCAAATCTAGATCAA 4338

Qy   1027 AAATCTCAACTTGTCTCGGTCTCCATGGTTTTGGAGAGTCTGATGATGAAGACGAA 1086
Db   4337 AAATCTCAACTTGTCTCGGTCTCCATGGTTTTGGAGAGTCTGATGATGAAGACGAA 4278

Qy   1087 GAAGATGAAAATTGATTGATCCATCTGAAAATTCATTCCGTCGTGCGACCACTTCGTCTT 1146
Db   4277 GAAGATGAAAATTGATTGATCCATCTGAAAATTCATTCCGTCGTGCGACCACTTCGTCTT 4218

Qy   1147 TCTTCCGGATTCGTTGAGAAATTAAAGTCAAATGAATTGAAAGTTGCATTGGACAGA 1206
Db   4217 TCTTCCGGATTCGTTGAGAAATTAAAGTCAAATGAATTGAAAGTTGCATTGGACAGA 4158

Qy   1207 ATTAAAATATCGAGTTGATGACGTGGAAAAGTATCTTGCTCCAAAGCCGATGAATTCAAT 1266
Db   4157 ATTAAAATATCGAGTTGATGACGTGGAAAAGTATCTTGCTCCAAAGCCGATGAATTCAAT 4098

Qy   1267 CCAAAACCTCAGCCTGGCTACTTTGCTCCACGTAAAATCCCAACAGACCACGTAAAATG 1326
Db   4097 CCAAAACCTCAGCCTGGCTACTTTGCTCCACGTAAAATCCCAACAGACCACGTAAAATG 4038

Qy   1327 CTTCCATTATTAATTGGATCTGATCCAAAGTTCAAGAGGAAATACGAAGACATCCAAGT 1386
Db   4037 CTTCCATTATTAATTGGATCTGATCCAAAGTTCAAGAGGAAATACGAAGACATCCAAGT 3978

Qy   1387 ACCGAAATGCAAAAGAATCAAGAATGTTTGACAAATTGAAGAATAATCCAAGT 1446
Db   3977 ACCGAAATGCAAAAGAATCAAGAATGTTTGACAAATTTGAAGAATAATCCAAGT 3918

Qy   1447 CTTGCTGCATTGTTCATCGATGATAAATTAGAGAATACATTGAAGGAGCAAATGTTA 1506
Db   3917 CTTGCTGCATTGTTCATGGATGATAAATTAGAGAATACATTGAAGGAGCAAATGTTA 3858

Qy   1507 ACTGATGAACAGAAAGGTAGAACACGTGTCAAACAATTCGTGCATTACCAAGACTGTTC 1566
Db   3857 ACTGATGAACAGAAAGGTAGAACACGTGTCAAAACAATTCGTGCATTACCAAGACTGTTC 3798

Qy   1567 GGTGCCACCAACTGCAAAGCTGAAATGATGATCGCAA-------------------- 1603
Db   3797 GGTGCCACCAACTGCAAAGCTGAAATGATGATCGCAAAGGTATAATGAGATTAATTATT 3738

Qy   1604 --------------AGGTATTCCAAGATATTGAAGAACGTCCCATTCCTCC 1640
Db   3737 TTGGCAATTTAAATATTATTCAGGTATTCCAAGATATTGAAGAACGTCCCATTCCTCC 3678

Qy   1641 ATTGTTCTTGAACCAAAGGAAGGCATACGAGATTGAGATGGACTGGAGCAAATGAAAA 1700
Db   3677 ATTGTTCTTGAACCAAAGGAAGGCATACGAGATTGAGATGGACTGGAGCAAATGAAAA 3618

Qy   1701 AGAAATTCCAGGACTT----------------------------------------- 1716
Db   3617 AGAAATTCCAGGACTTGGTAGGTATTGATTAAAAATGTTTTTGTCTTGGTTTTCTGCAAT 3558

Qy   1717 --------------------------------------------------------- 1716
Db   3557 GTCAGTGAAAACATAGAATGTGTACGAGAGAACTATGTTACGTGCCGACAACAGAAATTGA 3498

Qy   1717 ----------------------------------------GAAGTCGCTTCATTCTC 1734
Db   3497 TGAGTCAAAACAGAGAGAAAAACGCCAATCCTTTCTTATAGGAAGTCGCTTCATTCTC 3438

Qy   1735 CCATCTCTTGATCCAACTATGCCAGCCTTCTGAACACGGCTTTCTGACTCAGGGGCGAGCC 1794
Db   3437 CCATCTCTTGATCCAACTATGCCAGCCTTCTGAACACGGCTTTCTGACTCAGGGGCGAGCC 3378

Qy   1795 CGTGCAGCGAGTGGGATACCCATGTTCAAAATCCCGAATAACTGGAATCCTTGGAGATGAAGTT 1854
Db   3377 CGTGCAGCGAGTGGGATACCCATGTTCAAAATCCCGAATAACTGGAATCCTTGGAGATGAAGTT 3318

Qy   1855 GGGTTCAAAATGAACTCAAAAAACCAAACGATTCGTTGGAGGAAATGGAGCATTTGATATG 1914
Db   3317 GGGTTCAAAATGAACTCAAAAAACCAAACGATTCGTTGGAGGAAATGGAGCATTTGATATG 3258

Qy   1915 CCTGCACTGGGATTGTAG 1932
Db   3257 CCTGCACTGGGATTGTAG 3240
```

RESULT 3
CBRG03O12/c

```
LOCUS       CBRG03O12              17418 bp    DNA       linear   INV 04-NOV-2000
DEFINITION  Caenorhabditis briggsae cosmid G03O12, complete sequence.
ACCESSION   AC084485
VERSION     AC084485.1  GI:11094935
KEYWORDS    HTG.
SOURCE      Caenorhabditis briggsae
  ORGANISM  Caenorhabditis briggsae
            Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
            Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE   1  (bases 1 to 17418)
  AUTHORS   Washington University Genome Sequencing Center.
  TITLE     The C. briggsae Genome Sequencing Project
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 17418)
  AUTHORS   Wilson,R.
  TITLE     The sequence of C. briggsae cosmid G03O12
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 17418)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-NOV-2000) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT     Submitted by:
```

```
VERSION     AC084550.1  GI:11095000
KEYWORDS    HTG.
SOURCE      Caenorhabditis briggsae
  ORGANISM  Caenorhabditis briggsae
            Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
            Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE   1  (bases 1 to 39089)
  AUTHORS   Washington University Genome Sequencing Center
  TITLE     The C. briggsae Genome Sequencing Project
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 39089)
  AUTHORS   Wilson,R.
  TITLE     The sequence of C. briggsae cosmid G33D04
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 39089)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-NOV-2000) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, .St. Louis, Missouri 63108, USA
COMMENT     Submitted by:
            Genome Sequencing Center
            Department of Genetics, Washington University,
            St. Louis, MO 63110, USA
            e-mail: jspieth@watson.wustl.edu

            NOTICE: This sequence may not be the entire insert of this clone.
            It may be shorter because we only sequence overlapping sections
            once, or longer because we provide a small overlap between
            neighboring submissions.
FEATURES          Location/Qualifiers
     source       1..39089
                  /organism="Caenorhabditis briggsae"
                  /mol_type="genomic DNA"
                  /strain="GujArat G16"
                  /db_xref="taxon:6238"
                  /clone="G33D04"
ORIGIN

Query Match    10.9%;    Score 211;  DB 3;  Length 39089;
Best Local Similarity  77.3%;   Pred. No. 7.5e-38;
Matches 256;  Conservative  0;  Mismatches  75;  Indels  0;  Gaps  0;

Qy      40 TGTATCGCACCATTATCGGCCCAATCTCCATCGACTTCCGATGCTCCGGGAGCTTTGTTG 99
           ||||||||  ||| |  |||| |||||| |||||| |||||||| |||  |||||||
Db    1693 TTTCTAGATCCAATATCCGCGGGCTCACCACCGAATTCCGATGTTCCCGGTGCTCTTTTA 1634

Qy     100 TCATCTCTCGTAGGTAAAGCCATCAAAAACTACCACTGGCTCAATGACATCCAAGT 159
           || ||||| |  |||||  |  |||||| |||| ||  ||| ||  |  ||||| ||
Db    1633 TCCTCTCTTGCCACTAAAAACCACCCAAAACTTCCATTAGCTCCATCGAAGCACTA 1574

Qy     160 GAACTGATGGGTGTTCAATTTGTTGATGCTCTCATCAAAAAGGTCAAATGGAAATGCA 219
           ||| ||||||  ||||||||| ||||||||   || |||  |  ||| ||||||| ||
Db    1573 GAATTGATGGAGTTCAATTTGTGGATGCTTTTAATTAAAAAAAGGAAATGGAAATGCA 1514

Qy     220 AAAGGAGCATTTAAGACTCAATTAGAAGTTCTAGAGAAAGTACATCCTGATCAATTCGAT 279
           ||||||||||| ||  |  |||||  |  ||| |  ||||||||||| ||||||||  |
Db    1513 AAAGGAGCATTCAAAACAACACACTGGAGGTTCGGAAAGTACATCCGGATCAATTGTTGAG 1454

Qy     280 AAGTACAAAAGCTAAAAGTTGATGATTTGGCAGCTCGATCGCAGTTATGCAACAGGCGGAG 339
           ||  || |||   |||| || |||||||| ||| |||||  |||||||||||||||||||
Db    1453 AAATATAAGAAACTGAAAGTGGAGGATTTAGCGGCAGATGCAGTTATGCAACAGGCGGAA 1394

Qy     340 ATGGCAAAATTACAGCCTAAATCAGGAAATG 370
           |||  |||  || ||||   |  ||  |
Db    1393 ATGTCCAAGTTGCAACCGCAAAAACTGGTAAGG 1363
```

```
RESULT 5
I66494/c
LOCUS       I66494                    7218 bp    DNA     linear   PAT 28-DEC-1997
DEFINITION  Sequence 14 from patent US 5670367.
ACCESSION   I66494
```

```
Genome Sequencing Center
Department of Genetics, Washington University,
St. Louis, MO 63110, USA
e-mail: jspieth@watson.wustl.edu

NOTICE: This sequence may not be the entire insert of this clone.
It may be shorter because we only sequence overlapping sections
once, or longer because we provide a small overlap between
neighboring submissions.

FEATURES          Location/Qualifiers
     source       1..17418
                  /organism="Caenorhabditis briggsae"
                  /mol_type="genomic DNA"
                  /strain="GujArat G16"
                  /db_xref="taxon:6238"
                  /clone="G03012"
ORIGIN

Query Match    22.1%;    Score 426.4;  DB 3;  Length 17418;
Best Local Similarity  80.2%;   Pred. No. 1.5e-87;
Matches 515;  Conservative  0;  Mismatches 121;  Indels  6;  Gaps  1;

Qy     973 GATGAGGATGAAAGTGCATTCCGCGAATGGAGGCTCGTGCAAAACTAGATCAAAAATCT 1032
           ||||| ||||||| ||||| |||| | |||| |  ||||||| ||  ||||||||||
Db   14182 GATGAAGACGAAAACATATTCCGTGCGATGGAAGCCCGTGCAAAGCTCGATCAAAAATCG 14123

Qy    1033 CAACTTGTGCTCGGTCCCATGGTTTTG-----GAGAGTCGTGATGATGAAGACGAA 1086
           |||| | |  |||| ||| |    ||  |||| ||  |||| ||| |||||||
Db   14122 CAATTAGTTCTTCGGATTCCATGAATTCGATGAATGGGAGATGAAGATGAGGACGATGAG 14063

Qy    1087 GAAGATGAAAATTGATTGATCCATCTGAAAATTCATTCCGTCGCCACCACTTCGTCTT 1146
           ||||||||| || |||  |||||  ||   |||||||||| |||||||||  ||| ||
Db   14062 GAAGATGAGATTCAATGATCCATCGATCAGCCAACTCATTCCGTCGTCGTCCCACTTCGATTG 14003

Qy    1147 TCTTCCGGATTCGTTGAGAAATTAAAGTCAAATGATGAAAAGGTGCATTGGACAGA 1206
           |||||||||||||  | ||||  ||| ||||||||||| ||  |||||||||||||
Db   14002 TCTTCTGGATTCGTTGAAAAAACAAATGAATGAAATGAATGAGAACAGA 13943

Qy    1207 ATTAAATATGGAGTTGATGACGTGGAAAAGTATCTTGCTCCAAAGCCGATGGAATTCAAT 1266
           || ||||||  |||| | ||||||||||| |  ||||| ||||||||||| |||||||||
Db   13942 ATCAAGTATCGAGTTGACGATGTGGAAAAGTATTTGCAACCAAAGCCAAATGGATTTCAAT 13883

Qy    1267 CCAAAACCTCAGCCTGGCTACCTTGCTCGACAAGACCAACAAGACACGTAAAATG 1326
           || |||||| |||||||  | || |||| |||| |||||| || ||| ||||||
Db   13882 CCTAAACCCCGACCTGGATATTTGTACCTCGAAAAAATCCCGAAAAAATCCCGTAAAAATG 13823

Qy    1327 CTTCCATTATTAATTGGATCTGATCCAAAAGTTCAAGAGGAAAATACGAAGACATCCAAGT 1386
           ||||| |||| ||| || |||| ||||||||| |||||||| || |  || |||||||
Db   13822 CTTCCGTTGTTGATCGGCGACTCAGGACCTCAGAAGATATCCAAGC 13763

Qy    1387 ACGGATGGAAATTGCAAAGAATCAAGAGTTTTGAAGAATAATCCAAGT 1446
           || ||| ||| |||  ||||||||||| ||||||||||| ||||||||
Db   13762 AAAGAATGGAAGATTCAAGAGTTTTGAAGAATCAAGAGT 13703

Qy    1447 CTTGCTGCATTGTTCATGGATGAGAATAAATTGAAAGGAAGGCAAATGTTA 1506
           ||||||||| |  ||||| ||||| || ||  |||||||||||||||||
Db   13702 CTTGCTGCACTTTTTATGGGATGAGAATTGGAAAAATACGTTGGAAGAGGCAATTGTTG 13643

Qy    1507 ACTGATGAACAGAAAAGTAGAACACGTGTCAAAACAATTCGTGCATTACCAAGACTGTTC 1566
           |||||||| |||| ||| |   |||| || ||| |||| || ||| ||||||||||||
Db   13642 ACTGATGAGCAGAAGGAAGAAGAACCAGAGTTAAGACAATCCAGAGTTGCCCAGAGATTGTTT 13583

Qy    1567 GGTGCACCAACTGCAAAAGCTGAAAATGATTGATGCAAAGGTA 1608
           ||||||||||||| || | ||  |||||  |  ||||  |||
Db   13582 GGAGCACCAACTGCGAAAGCCGAGATGATCCGATGATCGCAAAGGTA 13541
```

```
RESULT 4
CBRG33D04/c
LOCUS       CBRG33D04              39089 bp    DNA     linear   INV 04-NOV-2000
DEFINITION  Caenorhabditis briggsae cosmid G33D04, complete sequence.
ACCESSION   AC084550
```

```
VERSION      I66494.1  GI:2724471
KEYWORDS     Unknown.
SOURCE       Unknown.
ORGANISM     Unclassified.
REFERENCE    1  (bases 1 to 7218)
AUTHORS      Dorner,F., Scheflinger,F. and Falkner,F.Gunter.
TITLE        Recombinant fowlpox virus
JOURNAL      Patent: US 5670367-A 14 23-SEP-1997;
FEATURES     Location/Qualifiers
source       1..7218
             /organism="unknown"
             /mol_type="unassigned DNA"
ORIGIN

Query Match         3.9%;  Score 75.4;  DB 6;  Length 7218;
Best Local Similarity  5.4%;  Pred. No. 1.6e-06;
Matches  22;  Conservative 237;  Mismatches 148;  Indels  0;  Gaps  0;

Qy    622 CAAGCTCTTGCTCAAAGCAATGCAATGTTAGGTAAAAATGCCCCAGTTGCCGGTGGAAGAGGT 681
          ||||| | ||||   |  |   |||   ||   ||| |||| ::: |||||| |||| |||
Db   1466 CAAGTAGTTAAAGAGATAGAAGAATTTCGTACRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1407

Qy    682 GAAGAACAACGGATGATGATGAGTGAGTGGACCAAAGAATGCAACAAAGAGAACTTCAA 741
          ::::: :::: ::: ::: ::: |::: |||  | ::: ::::::::::::::::::::
Db   1406 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1347

Qy    742 GAGGAAGATGAAGAGATGATGATCTTGAGGACGAGAGTGTACCCAGAAGAAGAAGTTCG 801
          ::::::: :::: : : :: :::  :: :::::::::::::::::::::: :::::::::
Db   1346 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1287

Qy    802 GATGGAGAACCACAAAGTGAAGCAGCAGAGCCATCAGAGAGAGATTTAGCCAGGAGATTGAAA 861
          :::::::: :::: :::: ::: :::::: :::: :::::::::: ::::::::::::::::::
Db   1286 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1227

Qy    862 AGTAGTCCTAGATTAAAGAGAGCTTTTACAGAATGCGGAAGTTCAATCATTGCTCTCTTAC 921
          :: ::: :::::: : :::::::::::::::::::::::::: :: ::::::::::::::::
Db   1226 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1167

Qy    922 CAACGAATGAGGGATTCTCCACTGAGCAAGGCGAAGGCCTTTGGCTATGAACGATGAGGAT 981
          ::::: :::::::: ::: ::::::::::::::: ::::::::::::::::::::::::::::
Db   1166 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1107

Qy    982 GAAAGTGCATTCCGCGCAATGGAGGCTCGTGTGCAAAAACTAGATCAAAA 1028
          ::::: ::::::::: ::: :::::::::::::::::::::::::::::::
Db   1106 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRATCGCAA 1060

RESULT 6
AX655393
LOCUS        AX655393         2000 bp    DNA    linear    PAT 22-MAR-2003
DEFINITION   Sequence 5263 from Patent WO03000898.
ACCESSION    AX655393
VERSION      AX655393.1  GI:29158207
KEYWORDS     .
SOURCE       Oryza sativa
ORGANISM     Oryza sativa
             Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
             Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
             Ehrhartoideae; Oryzeae; Oryza.
REFERENCE    1
AUTHORS      Chang,H.S., Chen,W., Cooper,B., Glazebrook,J., Goff,S.A., Hou,Y.M.,
             Katagiri,F., Quan,S., Tao,Y., Whitham,S., Xie,Z., Zhu,T. and Zou,G.
TITLE        Plant genes involved in defense against pathogens
JOURNAL      Patent: WO 03000898-A 5263 03-JAN-2003;
             Syngenta Participations AG (CH)
FEATURES     Location/Qualifiers
source       1..2000
             /organism="Oryza sativa"
             /mol_type="unassigned DNA"
             /db_xref="taxon:4530"
ORIGIN
```

```
Query Match          3.0%;  Score 57.2;  DB 6;  Length 2000;
Best Local Similarity  10.8%;  Pred. No. 0.026;
Matches  74;  Conservative 322;  Mismatches 280;  Indels  12;  Gaps  3;

Qy    726 ACAAAGAGAACTTCAAGACGAAGATGAAGATGATGATCTTCGAGGACGAGAGGATGTACC 785
          : :  ::: :::::::::: ::: ::: ::: ::: :::::|:||: :||::::::||::
Db    346 RMMMRSMCRWSKACYMRWRMWRMWRMWRMWTRRRRWAKKSSRTSRRKRKWCMRKRKYKRMRGYSR 405

Qy    786 CAGAAGAAGAAGTTCGGATGGAGAACCACCAAAGTGAAGCAGCATCAGAGAAGAGATTT 845
          ::: :|.|: :|:|::|| | |::::::::::: :|: : :::: ::::: ::::::::
Db    406 MRSCKRARWMKRCRSGRAWKMGCRGCMTCRMKSYGMRWKSWKRMASKYKWMSRMYRWRK 465

Qy    846 AGCCAGGAGATTGAAAAGTAGTCCTAGATTAAAAGAGCTTTTACAGAATGCGGAAGTTCA 905
          : ||: :: : :::::||| : :::: :||::::::::::: :::: :::::::::|:||
Db    466 KKCSRTTMWGKTRGGMGTMGRCRYKKRSGMKRKCRRRWGRMYRMRWKRYYMSARYTMR 525

Qy    906 ATCATTGCTCTCTTACCAACGAATGAGGGATTCTCCACTGAGCAAGGCCTTTGGC 965
          : |: :|| ||||| :::::::: ::::::: ::::::::::::::  :::::::
Db    526 YCARKKKYSYSSAARKARCWYRGKGYYWAGMMKRYKRMYMKMMWYKRKYSKCSWYCKMS 585

Qy    966 TATGAACGATGAGGATGAAGTGCATTCCGCCGCAATGGAGGCTCGTGCAAAACTAGATCA 1025
          : ::: :| ||| |:|:|:::::::::::::::::::::: ::|::::|:|:|
Db    586 YYASCMKSARKAGAKMCKRSKSMSAWSKSMRSSRKCRKCASKRSSAKRYAMMG--GMTSG 642

Qy   1026 AAAATCTCAACTTGTCTCGGTCTCCATCGTTTGGAGAGTCTGATGATGAAGACGA 1085
          :: :::::::::|||| | :| || ::::::::::::::: :::------RYRAWCMYMW 695
Db    643 SRMSRWKSYTCYWRKWGSMKSTCTWMYYMSKYTYAKYGSYWRY------RYRAWCMYMW 695

Qy   1086 AGAAGATGAAAATTGATTGATCATCTGAAAATTCATTCCGTCGTGCACCACTTCGTCT 1145
          ::::::::::::::::::::|:: ::: ::::::::::: ::: :::::::::::::::
Db    696 RWYYYRYRSYMTYMAWYTSSTRRAMTGMKYSGRYWTSMWYKYCKCSWKYRSMWYYWSWWNA 755

Qy   1146 TTCTTCCGGATTCGTTGGAGAAATTAAAGTCAAATGATGAATTGAAAAGTGCATTGGACAG 1205
          ::|:| ::| :: : :::::::::::: ::::::::::::::::::::::::::::: ::|
Db    756 KTWKMWRRYATRMWMWMYRYSMKWYTWCTMWGYWVYWRTYMKMRYMWYKCTKTYWYWSA 815

Qy   1206 AATTAAATATCGAGTTGATGACGTGGAAAAGTATCTTGCTCCAAAGCCCGATGGAATTCAA 1265
          ::|::: ::: :::: ::: |::: ::::::::::: ::::::::::::::::::::::::
Db    816 TYWTGTWAAWWMAKTKWRMGMTGAKTRGRARKARYWWKWATWCATKRWMTKGKGAKWAWT 875

Qy   1266 TCCAAAACCTCAGCCTGGCTACTTTGCTCCACGTAAAATCCCAACAACAGACCACGTAAAAT 1325
          ::: ::: ::::::::::::: ::|::::::: :::::: :::: ::: :::::::::::::
Db    876 WMAKAWRKYWSWM--RAWYYYKTRRTRYKTCWWKARWGSWAYWRMWWKGSAKWMWWWK 933

Qy   1326 GCTTCCATTATTAATTGGATCTGATCCAAAAGTTCAAGAGGAAAATACGAAGACATCCAAG 1385
          | ::: :::::| |::||||:::::|:::|:: :::::::|::::::::|::: :::: :::
Db    934 GGRWGWTKYWYWYCTTWKMACGRATKYMCCAGWWAMYSYSWTRTYWMRTWRWMWASSRTA 993

Qy   1386 TACCGAATGGAAAATTGCAAAAGAATCA 1413
          ::: ::|::::::::|::::::|:: ||:||
Db    994 KRMARMWMWKTRAWSKSYARAYWKMAGCA 1021
```

```
RESULT 7
AC111055
LOCUS        AC111055        161964 bp    DNA    linear    HTG 23-MAR-2004
DEFINITION   Mus musculus chromosome 10 clone RP24-502H14 map 10, *** SEQUENCING
             IN PROGRESS ***, 9 ordered pieces.
ACCESSION    AC111055
VERSION      AC111055.6  GI:45642845
KEYWORDS     HTG; HTGS_PHASE2; HTGS_FULLTOP; HTGS_ACTIVEFIN.
SOURCE       Mus musculus (house mouse)
ORGANISM     Mus musculus
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE    1  (bases 1 to 161964)
AUTHORS      Birren,B., Nusbaum,C. and Lander,E.
TITLE
JOURNAL      Unpublished
REFERENCE    2  (bases 1 to 161964)
AUTHORS      Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
             Anderson,S., Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B.,
             Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
```

```
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
            Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
            Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R.,
            Landers,T., Lehoczky,J., Levine,R., Liu,G., MacLean,C.,
            Macdonald,P., Major,J., Marquis,N., Matthews,C., McCarthy,M.,
            McEwan,P., McKernan,K., Meldrim,J., Meneus,L., Mihova,T.,
            Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C.,
            Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J.,
            Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C.,
            Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J.,
            Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S.,
            Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,J., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
TITLE       Direct Submission
JOURNAL     Submitted (18-FEB-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 161964)
AUTHORS     Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,
            Anderson,M., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T.,
            Boguslavkiy,L., Boukhgalter,B., Camarata,J., Chang,J., Choepel,Y.,
            Collymore,A., Cook,A., Cooke,P., Corum,B., DeArellano,K.,
            Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S.,
            Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S.,
            Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B.,
            Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
            Lindblad-Toh,K., Liu,X., Lui,A., Mabbitt,R., MacLean,C.,
            Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M.,
            Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J.,
            Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
            Rachupka,A., Ramaaamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P.,
            Roman,J., Schauer,S., Schupback,R., Seaman,S., Severy,P., Smith,C.,
            Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M.,
            Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M.,
            Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X.,
            Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
TITLE       Direct Submission
JOURNAL     Submitted (23-MAR-2004) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Mar 23, 2004 this sequence version replaced gi:44886783.
            All repeats were identified using RepeatMasker:
            Smit,A.F.A. & Green,P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            --------------- Genome Center
            Center: Whitehead Institute/ MIT Center for Genome Research
            Center code: WIBR
            Web site: http://www-seq.wi.mit.edu
            Contact: sequence_submissions@genome.wi.mit.edu
            --------------- Project Information
            Center project name: L20154
            Center clone name: 502_H_14
            --------------
          * NOTE: This is a 'working draft' sequence. It currently
          * consists of 9 contigs. Gaps between the contigs
          * are represented as runs of N. The order of the pieces
          * is believed to be correct as given, however the sizes
          * of the gaps between them are based on estimates that have
          * provided by the submittor.
          * This sequence will be replaced
          * by the finished sequence as soon as it is available and
          * the accession number will be preserved.
          *      1   6498: contig of 6498 bp in length
          *   6499   6598: gap of 100 bp
          *   6599  14424: contig of 7826 bp in length
          *  14425  14524: gap of 100 bp
          *  14525  35310: contig of 20786 bp in length
          *  35311  35410: gap of 100 bp
          *  35411  58107: contig of 22697 bp in length
          *  58108  58207: gap of 100 bp
          *  58208  74299: contig of 16092 bp in length
          *  74300  74399: gap of 100 bp
          *  74400 103424: contig of 29025 bp in length
          * 103425 103524: gap of 100 bp
          * 103525 156902: contig of 53378 bp in length
          * 156903 157002: gap of 100 bp
          * 157003 159646: contig of 2644 bp in length
          * 159647 159746: gap of 100 bp
          * 159747 161964: contig of 2218 bp in length.
FEATURES             Location/Qualifiers
     source          1..161964
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
                     /chromosome="10"
                     /map="10"
                     /clone="RP24-502H14"
                     /clone_lib="RPCI-24 Male Mouse BAC"

ORIGIN

Query Match             2.9%;  Score 56.2;  DB 2;  Length 161964;
  Best Local Similarity   49.8%;  Pred. No. 0.037;
  Matches 142;  Conservative  0;  Mismatches 143;  Indels  0;  Gaps  0;

Qy    635 AAAGAGCAATGTTAGTGAAAATGCCCAGTTGCCCGTGGAAGAGGTGAAGAACACGGA 694
          ||||||  ||| ||||| |||||| ||||||||||||||  |||||||||||
Db  54954 AAAGAAAAAAAACGAAGAAGAAGAAGAAGAAGAAGGAAGAAGAGAAGATGAAG 55013
Qy    695 TGATGATGAATCGAGTGGACCAAAGAATGCAACAAAGAACTTCAAGAGAGATGAAG 754
          |||  ||||||||||||||||  |||||||| ||
Db  55014 AAGAGGAAGACGAAGAGGAAGAGGAAGAGGAAGAGGAAGAGAAGAGAAGAGAAG 55073
Qy    755 ATGATGATGATCTTGAGGACGAAGAGAAGAAGTTCCGATGGAGAACCAC 814
          ||  |||||||||||||||  ||||||
Db  55074 AAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAGGAGGAGGAAG 55133
Qy    815 AAAGTGAAGCAGACAGCATCAGAGAAGAGATTTAGCCAGGAAGAAGTTGAAAAGTAGTCCTAGAT 874
          |||  ||||||||
Db  55134 AGGAAGGAGGAGGAAGGAGGAAGAGGAAGAAAAGGAAAAAGAAAAAAAATTCGTGTT 55193
Qy    875 TAAAAGAGCTTTTACAGAATGCGGAAGTTCAATCATTGCTCTCTT 919
          ||||||||||||
Db  55194 GAAGGGATTTTTTAAAAGTAGCAACTTATATTAATATATCATTT 55238

RESULT 8
AC147563/c
LOCUS       AC147563            236428 bp    DNA     linear   ROD 29-MAY-2004
DEFINITION  Mus musculus BAC clone RP23-231H21 from chromosome 10, complete
            sequence.
ACCESSION   AC147563
VERSION     AC147563.4  GI:47777619
KEYWORDS    HTG.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 236428)
  AUTHORS   Levy,A., Kozlowicz,A. and Meyer,R.
  TITLE     The sequence of Mus musculus BAC clone RP23-231H21
  JOURNAL   Unpublished (2001)
REFERENCE   2  (bases 1 to 236428)
  AUTHORS   Wilson,R.K.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-DEC-2003) Genome Sequencing Center, 4444 Forest Park
            Parkway, St. Louis, MO 63108, USA
REFERENCE   3  (bases 1 to 236428)
  AUTHORS   Wilson,R.K.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-MAY-2004) Genome Sequencing Center, 4444 Forest Park
```

```
REFERENCE   4  (bases 1 to 236428)
AUTHORS     Wilson,R.K.
TITLE       Direct Submission
JOURNAL     Submitted (28-MAY-2004) Genome Sequencing Center, 4444 Forest Park
            Parkway, St. Louis, MO 63108, USA
REFERENCE   5  (bases 1 to 236428)
AUTHORS     Wilson,R.K.
TITLE       Direct Submission
JOURNAL     Submitted (29-MAY-2004) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT     On May 28, 2004 this sequence version replaced gi:47131324.
            ---------------- Genome Center
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site: http://genome.wustl.edu
            Contact: submissions@watson.wustl.edu
            ---------------- Summary Statistics
            Center project name: M_BA0231H21
            ----------------

            NOTICE:

            This sequence was finished as follows unless otherwise noted:
            all regions were double stranded, sequenced with an alternate
            chemistry, or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by sequence
            from more than one subclone; and the assembly was confirmed by
            restriction digest.

            MAPPING INFORMATION:
            Mapping information for this clone was provided by Dr. Wes Warren,
            Department of Genetics, Washington University, St. Louis MO. For
            additional information about the map position of this sequence, see
            http://genome.wustl.edu

            SOURCE INFORMATION:
            The RPCI-23 BAC Library has been constructed by Kazutoyo Osegawa
            and Minako Tateno in the laboratory of Pieter de Jong
            (http://www.chori.org) from female C57BL/6J mouse kidney and/or
            brain genomic DNA.  The clone and detailed information can be
            obtained from Research Genetics, Inc. (http://www.resgen.com) or
            Pieter de Jong and coworkers at http://www.chori.org

            NEIGHBORING SEQUENCE INFORMATION:
            This sequence is the entire insert of the clone.
FEATURES             Location/Qualifiers
     source          1..236428
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
                     /chromosome="10"
                     /map="10"
                     /clone="RP23-231H21"
                     /clone_lib="RPCI-23"
     repeat_region   1..171
                     /rpt_family="MaLR"
     repeat_region   1561..1790
                     /rpt_family="B4"
     repeat_region   2037..2200
                     /rpt_family="B4"
     repeat_region   3815..3958
                     /rpt_family="Alu"
     repeat_region   4188..4637
                     /rpt_family="L1"
     repeat_region   4635..5812
                     /rpt_family="L1"
     repeat_region   5813..5929
                     /rpt_family="Alu"
     repeat_region   5986..6178
                     /rpt_family="L1"
     repeat_region   6175..6492
                     /rpt_family="L1"
     repeat_region   7740..7813
                     /rpt_family="L1"
     repeat_region   8115..8183
                     /rpt_family="Alu"
     repeat_region   9381..9515
                     /rpt_family="Alu"
     repeat_region   11530..12350
                     /rpt_family="L1"
     repeat_region   12351..12497
                     /rpt_family="Alu"
     repeat_region   12498..12754
                     /rpt_family="L1"
     repeat_region   12755..13176
                     /rpt_family="ERVK"
     repeat_region   13177..13360
                     /rpt_family="ERVK"
     repeat_region   13361..13734
                     /rpt_family="ERVK"
     repeat_region   13740..13974
                     /rpt_family="ERVK"
     repeat_region   13951..13986
                     /rpt_family="ERVK"
     repeat_region   13975..14058
                     /rpt_family="ERVK"
     repeat_region   14151..14443
                     /rpt_family="ERVK"
     repeat_region   14467..14864
                     /rpt_family="ERVK"
     repeat_region   14865..15350
                     /rpt_family="L1"
     misc_feature    21401..22231
                     /note="CpG_island (%GC=63.5, o/e=0.75, #CpGs=55)"
     repeat_region   22352..24002
                     /rpt_family="L1"
     repeat_region   24099..24536
                     /rpt_family="L1"
     repeat_region   24537..24935
                     /rpt_family="MaLR"
     repeat_region   24936..25626
                     /rpt_family="L1"
     repeat_region   25972..26123
                     /rpt_family="MER1_type"
     unsure          26141..26511
                     /note="Unresolved simple sequence repeat."
     repeat_region   26515..26698
                     /rpt_family="B4"
     repeat_region   27153..27395
                     /rpt_family="L2"
     repeat_region   29272..29412
                     /rpt_family="Alu"
     repeat_region   30330..30671
                     /rpt_family="MaLR"
     repeat_region   31294..31651
                     /rpt_family="MaLR"
     repeat_region   31930..32071
                     /rpt_family="Alu"
     repeat_region   34100..34153
                     /rpt_family="L1"
     repeat_region   34263..34997
                     /rpt_family="L1"
     repeat_region   34990..35758
                     /rpt_family="L1"
     repeat_region   36065..36203
                     /rpt_family="B4"
     repeat_region   37114..37261
                     /rpt_family="MER1_type"
     repeat_region   37363..37416
                     /rpt_family="MER1_type"
     repeat_region   37667..38886
                     /rpt_family="L1"
     repeat_region   39025..39212
```

```
tRNA            complement(39136..39207)
                /product="tRNA-Ser"
                /note="Likely pseudogene (HMM Sc=31.53 / Sec struct
                Sc=-10.51)"
repeat_region   41703..41871   /rpt_family="B2"
repeat_region   41905..42096   /rpt_family="L1"
repeat_region   43371..43469   /rpt_family="B2"
repeat_region   45482..45576   /rpt_family="ERVL"
repeat_region   45577..45594   /rpt_family="Alu"
repeat_region   45648..45793   /rpt_family="B4"
repeat_region   47666..47871   /rpt_family="Alu"
repeat_region   47907..47995   /rpt_family="MaLR"
repeat_region   48419..48564   /rpt_family="L1"
repeat_region   48750..48858   /rpt_family="MaLR"
repeat_region   49101..49272   /rpt_family="Alu"
repeat_region   49922..50229   /rpt_family="B2"
repeat_region   50992..51294   /rpt_family="MaLR"
```

```
Query Match      2.9%;   Score 56;  DB 10;   Length 236428;
Best Local Similarity  56.4%;   Pred. No. 0.041;
Matches  124;  Conservative  0;  Mismatches  95;  Indels  1;  Gaps  1;

Qy    663 AGTTGCCGGTGGAAGAGGTGAAGAACCACGATGATGAATCGAGTGACCAAAGAAT 722
              ||||||||| ||||||| |||||
Db  26366 AGGAGGAGGAAGGAGGAGGAAGAAGAAGAAGAAGAAGAAGAAGA-GAAGAAGAAGA 26308

Qy    723 GCAACAAAGAGAGAACTTCAAGAGGAAGATGAAGATGAAGATGATGATCTTGAGGACGAGGATGT 782
              ||| ||||||||| ||| ||||||| |||||||| |||||
Db  26307 AGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGACATCAGAGAAGAAGA 26248

Qy    783 ACCCAGAAGAAGAAGAAGTTCGGATGGAGAACCACAAAGTGAAGCAGGAGCATCAGAGAAGAAGAGA 842
              ||||||||||||| |||||||||
Db  26247 AGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAGGAGA 26188

Qy    843 TTTAGCCAGGAGATTGAAAAGTAGTCCTAGATTAAAAGAG 882
              ||||
Db  26187 AGAAGAAGAAGAAAGAAGAAGAAGAAGAGAGAGAGAGAGAGAG 26148
```

RESULT 9
AC116690
```
LOCUS       AC116690             171583 bp    DNA     linear   ROD 15-DEC-2003
DEFINITION  Mus musculus chromosome 8, clone RP24-279I6, complete sequence.
ACCESSION   AC116690
VERSION     AC116690.15  GI:39841452
KEYWORDS    HTG.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 171583)
  AUTHORS   Birren,B., Nusbaum,C. and Lander,E.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 171583)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L.,
            Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,
            Chazaro,L., Choepel,Y., Colangelo,M., Collins,S., Collymore,A.,
```

```
            Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
            Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R.,
            Landers,T., Lehoczky,J., Levine,R., Lindblad-Toh,K., Liu,G.,
            MacLean,C., Macdonald,P., Major,J., Marquis,J., Meneus,L.,
            McCarthy,M., McEwan,P., McKernan,K., Meldrim,J., Meneus,L.,
            Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R.,
            Norbu,C., Norman,C.H., O'Connor,T., O'Neil,D.,
            Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
            Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
            Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback.R.,
            Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic.N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-APR-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 171583)
  AUTHORS   Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,
            Anderson,M., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T.,
            Boguslavkiy,L., Boukhgalter,B., Camarata,J., Chang,J., Choepel,Y.,
            Collymore,A., Cook,A., Cooke,P., Corum,B., DeArellano,K.,
            Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S.,
            Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S.,
            Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B.,
            Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
            Lindblad-Toh,K., Liu,G., Lui,X., Mabbitt,R., MacLean,C.,
            Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M.,
            Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J.,
            Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
            Rachupka,A., Ramasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P.,
            Roman,J., Schauer,S., Schupback,R., Seaman,S., Severy,P., Smith,C.,
            Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M.,
            Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M.,
            Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X.,
            Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-NOV-2003) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   4  (bases 1 to 171583)
  AUTHORS   Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,
            Anderson,M., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T.,
            Boguslavkiy,L., Boukhgalter,B., Camarata,J., Chang,J., Choepel,Y.,
            Collymore,A., Cook,A., Cooke,P., Corum,B., DeArellano,K.,
            Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S.,
            Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S.,
            Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B.,
            Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
            Lindblad-Toh,K., Liu,G., Lui,X., Mabbitt,R., MacLean,C.,
            Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M.,
            Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J.,
            Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
            Rachupka,A., Ramasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P.,
            Roman,J., Schauer,S., Schupback,R., Seaman,S., Severy,P., Smith,C.,
            Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M.,
            Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M.,
            Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X.,
            Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-NOV-2003) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   5  (bases 1 to 171583)
  AUTHORS   Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,
            Anderson,M., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T.,
            Boguslavkiy,L., Boukhgalter,B., Camarata,J., Chang,J., Choepel,Y.,
```

```
           Collymore,A., Cook,A., Cooke,P., Corum,B., DeArellano,K.,
           Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S.,
           Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S.,
           Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B.,
           Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
           Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
           Lindblad-Toh,K., Liu,X., Lui,A., Mabbitt,R., MacLean,C.,
           Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M.,
           Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J.,
           Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P.,
           O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
           Rachupka,A., Ramasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P.,
           Roman,J., Schauer,S., Schupback,R., Seaman,S., Severy,P., Smith,C.,
           Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M.,
           Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M.,
           Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X.,
           Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
TITLE      Direct Submission
JOURNAL    Submitted (15-DEC-2003) Whitehead Institute/MIT Center for Genome
           Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT    On Dec 15, 2003 this sequence version replaced gi:38490603.
           All repeats were identified using RepeatMasker.
           Smit, A.F.A. & Green, P. (1996-1997)
           http://ftp.genome.washington.edu/RM/RepeatMasker.html
           ----------- Genome Center
               Center: Whitehead Institute/ MIT Center for Genome Research
               Center code: WIBR
               Web site: http://www-seq.wi.mit.edu
               Contact: sequence submissions@genome.wi.mit.edu
           ----------- Project Information
               Center project name: L21083
               Center clone name: 279_I_6
           -----------
FEATURES             Location/Qualifiers
     source          1..171583
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
                     /chromosome="8"
                     /map="8"
                     /clone="RP24-27916"
                     /clone lib="RPCI-24 Male Mouse BAC"
     repeat_region   complement(1..508)
                     /rpt_family="L1F"
     misc_feature    1..125
                     /note="wgs end extension
                     clone end:SP6"
     unsure          58..62
                     /note="<30 qual SNGL region"
     misc_feature    126..129
                     /note="clone boundary
                     clone end:SP6
                     site:MboI"
     repeat_region   1366..1699
                     /rpt_family="MTE"
     repeat_region   2128..2148
                     /rpt_family="AT_rich"
     repeat_region   2554..2582
                     /rpt_family="AT_rich"
     repeat_region   complement(2785..3360)
                     /rpt_family="L1MC3"
     repeat_region   complement(3397..3817)
                     /rpt_family="L1MC3"
     repeat_region   4299..4526
                     /rpt_family="RMER20"
     repeat_region   4968..5012
                     /rpt_family="AT_rich"
     repeat_region   complement(5602..5787)
                     /rpt_family="B3A"
     repeat_region   8269..8323
                     /rpt_family="(TG)n"
     repeat_region   8373..8399
                     /rpt_family="AT_rich"
```

```
     repeat_region   complement(8401..8563)
                     /rpt_family="B4"
     repeat_region   8691..8747
                     /rpt_family="(TTTG)n"
     repeat_region   10704..10752
                     /rpt_family="Lx9"
     repeat_region   10773..10799
                     /rpt_family="(CA)n"
     repeat_region   11176..11252
                     /rpt_family="(TCTA)n"
     repeat_region   11403..11544
                     /rpt_family="(TAGA)n"
     repeat_region   11561..11713
                     /rpt_family="(TAGA)n"
     repeat_region   11727..11782
                     /rpt_family="(TAGA)n"
     repeat_region   13879..13941
                     /rpt_family="(A)n"
     repeat_region   13990..14011
                     /rpt_family="(CAAAAA)n"
     repeat_region   14442..14510
                     /rpt_family="(TG)n"
     repeat_region   15071..15205
                     /rpt_family="L1MA9"
     repeat_region   15206..15239
                     /rpt_family="(CA)n"
     repeat_region   15240..15832
                     /rpt_family="L1MA9"
     repeat_region   15833..15874
                     /rpt_family="L1MA9"
     repeat_region   15875..16103
                     /rpt_family="(CA)n"
     repeat_region   16104..16198
                     /rpt_family="L1MA9"
     repeat_region   16199..16496
                     /rpt_family="L1MA9"
```

```
Query Match          2.9%;  Score 55.4;  DB 10;  Length 171583;
Best Local Similarity  55.5%;  Pred. No. 0.057;
Matches 127;  Conservative  0;  Mismatches 101;  Indels  1;  Gaps  1;

Qy      653 AAAATGCCCAGTTGCCGGTGGAAGAGGTGAAGAACAACGGATGATGAATCGAGTGG 712
            |||||| ||||| ||| |||||||||| |||||||||| |||| | |||||| |||
Db   158455 AAGAAGAAGAAGAAGGAGGAGGAGGAGGAGGAGGAGGAAGATGAAGATGATGGAAGAAGAAG 158514

Qy      713 ACCAAAGAATGCAACAAAGAGAACTTCAAGAGGAAGAAGATGAAGATGATGATGGGG 772
            |||| ||||   |||| |||||| ||||||||||||||||||||||||||||||||
Db   158515 AGGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAG 158574

Qy      773 ACGAGGATGTACCCAGAAGAAGAAG-TTCGGATGGAGAACCACAAAGTGAAGCAGAGCAT 831
            |||||||||   |||||||||||||  ||||||||||||||||||||||||||||||||
Db   158575 AAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAG 158634

Qy      832 CAGAGAAAGAGATTTAGCCAGGAGGATTGAAAAAGTAGTAGTCCTAGATTAAAAG 880
            |||||||||||   |||||||||| ||||||||||||||| |||||||| |||||
Db   158635 GAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAG 158683

RESULT 10
BX842643
LOCUS       BX842643            346939 bp    DNA     linear   BCT 10-JUN-2004
DEFINITION  Mycoplasma mycoides subsp. mycoides SC genomic DNA, complete
            sequence; segment 2/4.
ACCESSION   BX842643 BX293980
VERSION     BX842643.1  GI:42492348
KEYWORDS    .
SOURCE      Mycoplasma mycoides subsp. mycoides SC
  ORGANISM  Mycoplasma mycoides subsp. mycoides SC
            Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
REFERENCE   1
  AUTHORS   Westberg,J., Persson,A., Holmberg,A., Goesmann,A., Lundeberg,J.,
            Johansson,K.E., Pettersson,B. and Uhlen,M.
  TITLE     The genome sequence of Mycoplasma mycoides subsp. mycoides SC type
```

```
JOURNAL          strain PG1T, the causative agent of contagious bovine
                 pleuropneumonia (CBPP)
PUBMED           14762060
REFERENCE        2  (bases 1 to 346939)
AUTHORS          Westberg,J.
TITLE            Direct Submission
JOURNAL          Submitted (07-MAR-2003) Westberg J., Royal Institute of Technology
                 (KTH), Department of Biotechnology, Roslagstullsbacken 21, SE-106
                 91 Stockholm, Sweden (E-mail: jockew@biotech.kth.se)
FEATURES             Location/Qualifiers
     source          1..346939
                     /organism="Mycoplasma mycoides subsp. mycoides SC"
                     /mol_type="genomic DNA"
                     /strain="PG1"
                     /specific_host="Cattle and buffalo"
                     /db_xref="taxon:44101"
     gene            complement(86..754)
                     /locus_tag="MSC_0304"
     CDS             complement(86..754)
                     /locus_tag="MSC_0304"
                     /codon_start=1
                     /evidence=not_experimental
                     /transl_table=4
                     /product="CONSERVED HYPOTHETICAL PROTEIN"
                     /protein_id="CAE76945.1"
                     /db_xref="GI:42492349"
                     /translation="MVVVILQKEQKDYLMYKTILIVTAKTNLNLEIFNNDSTYVIGV
                     ERGCLDLISRNMKIDYAIGDFDKVTDKELELIKSKIKNVEIWSPEKDYFDGELAILKG
                     LKASKDAKIFIANATKRYDKNYSIFHFIFKYDLTFINDDSILFRFDKGTNILKFEDY
                     QDYTYVSFIAKDNTKITIKNLKYRCENLSLSAYSNSCLSNAFLPYKDGIIWSNNPIIC
                     ILTK"
     gene            complement(702..1379)
                     /locus_tag="MSC_0305"
                     /gene="rpe"
     CDS             complement(702..1379)
                     /locus_tag="MSC_0305"
                     /gene="rpe"
                     /EC_number="5.1.3.1"
                     /codon_start=1
                     /evidence=not_experimental
                     /transl_table=4
                     /product="ribulose-phosphate 3-epimerase"
                     /protein_id="CAE76946.1"
                     /db_xref="GI:42492350"
                     /translation="MKKYIIAPSVLSANFMDLKNELELCKKNNINWHIYDVMDFDFVP
                     NLTFGSKILHDIKKNIDINVDVHFMVSVKTKQFEDFFSDYIKAKPEMWTMHIESLKDN
                     LEKIQILDQLRKERKYKYIIEVDGGINEQTSVLVKQAGVDMIVAGSYLFGSGDFTKRA
                     KGLFDEL"
     gene            complement(1381..2283)
                     /locus_tag="MSC_0306"
     CDS             complement(1381..2283)
                     /locus_tag="MSC_0306"
                     /codon_start=1
                     /evidence=not_experimental
                     /transl_table=4
                     /product="Conserved hypothetical protein"
                     /protein_id="CAE76947.1"
                     /db_xref="GI:42492351"
                     /translation="MQGILIKKDSNESYVFYKNNIYKVYAKGNLKKELKPKVGDIVEF
                     SILEDGNGYIKEVKPRKNSLDRPKIANVDQVLIVTALYEPLFSSFLLNKYIMMLETLK
                     IKPILIFTKVDLLKQTQYQEVMHKINWYKKMHYQVLIINNKDLENKKEIILKLKEL
                     LKNKISVFTGQTGAGKSTTLNFLDINSQIKTNEISKKLNRGKHTTTSIQLYNLENDI
                     FIADTPGFSAFDLNNIDIEHILYSWETFIPFLNKCKFVDCTHTHETKCEVKKAVAEHL
                     LPTFIYNDYVKVYEELKQNKKNKY"
     gene            complement(2293..3408)
                     /gene="pkn"
                     /locus_tag="MSC_0307"
     CDS             complement(2293..3408)
                     /gene="pkn"
                     /locus_tag="MSC_0307"
                     /EC_number="2.7.1.-"
                     /codon_start=1
                     /evidence=not_experimental
                     /transl_table=4
                     /product="serine/threonine protein kinase"
                     /protein_id="CAE76948.1"
                     /db_xref="GI:42492352"
                     /translation="MPKTKDLRINKEELLNQVVNRYKLIKYLNSGASAVVFKALDL
                     DASVLEKKDVFIAVKIILKAKNKNIETIKKCRLFLETNIFAKLSFSKNIVKMKDVFSWQ
                     NYYVIVMELIEGVDLNKKFNAYNNVLSNKEFIYYFLEITKGLKEIHDNNIIHKDVKPA
                     NILITNDSKVRISDFGISIKSIILDDHHNHISPGTPRYTAPEQFINFESRKDALYFE
                     SDIYSTGVIMYEFLTGSMLYLNYGSNHTSSKEKELTNFQQHILKDITRPEINPNISQ
                     ALENIIMKCLAKDYKNRYHTFDQIIKDLEQAKQQPDVNIDFPNMWWEDENVLNIKONN
                     TLKYKYFFKNTNFKYFLFWISIVISLFIIFLIVLLIK"
     gene            complement(3727..4152)
                     /locus_tag="MSC_0308"
     CDS             complement(3727..4152)
                     /locus_tag="MSC_0308"
                     /EC_number="3.1.3.16"
                     /note="truncated"
                     /codon_start=1
                     /evidence=not_experimental
                     /transl_table=4
                     /product="protein phosphatase"
                     /protein_id="CAE76949.1"
                     /db_xref="GI:42492353"
                     /translation="MKIKAVGLSEKGNFRKENQDFLNYYKNSNGSFLAIICDGMGGHK
                     HGEIASRLAVNTLVDLFKKTTFDKKESKEIFSWLEKSVSYIINVMKKYVEDFIEAKDM
                     GTTLSAILIANNNAHIITNLSFKRSKFKSNYNWSKFNEF"
     gene            complement(4366..4551)
                     /locus_tag="MSC_0309"
     CDS             complement(4366..4551)
                     /locus_tag="MSC_0309"
                     /codon_start=1
                     /evidence=not_experimental
                     /transl_table=4
                     /product="Hypothetical protein"
                     /protein_id="CAE76950.1"
                     /db_xref="GI:42492354"
                     /translation="MFLTAFDTEVKDNILIIKPKKEYEYFYNQKQILLKPKDENINT
                     INSDTNNQNKLTNSNIN"
     gene            complement(4825..5598)
                     /locus_tag="MSC_0310"
     CDS             complement(4825..5598)
                     /locus_tag="MSC_0310"
                     /codon_start=1
                     /evidence=not_experimental
                     /transl_table=4
                     /product="Conserved hypothetical protein"
                     /protein_id="CAE76951.1"
                     /db_xref="GI:42492355"
                     /translation="MIKMFLILLSSMNLTNIKEPNLIKTNIIQNYRSEFNLSEKVNIS
                     RTIKASNTPSKNNDNNFYILKSSANFYNQTVETILKGFNFKLSDEYISFIKDGFKLEI
                     SLTVNLEFINTKDQYFNNKYDLVFYLSNNSFTNQSSKDIIWKNVRYQTSDRMITDYID
                     TKNHWQFSFNKELYLFNLKHVFNKKYNFRFTDFNNLYSFNSLKYNIDTIKYRFYTQLS
                     LYSQIKDFNFNINKDNISVNDIKKQYIQQTNLKNSEIL"
     gene            5697..6125
                     /locus_tag="MSC_0311"
     CDS             5697..6125
                     /locus_tag="MSC_0311"
                     /codon_start=1
                     /evidence=not_experimental
                     /transl_table=4
                     /product="Conserved hypothetical protein"
                     /protein_id="CAE76952.1"
                     /db_xref="GI:42492356"
                     /translation="MRLYMKVKVNNLEQTYKIAKKIKKIIQNKIPFYILLKGDLGAG
                     KTTFTKALLEQFNIKDNIISPSFVIMNQYFIDDLKINHMDAYRLNNDSEIEMYLDEFL
                     DGLNIIEWYENLDLDLNAINKLIIEIKIIDENQRLFFIGE"
     gene            6127..6690
                     /locus_tag="MSC_0312"
     CDS             6127..6690
                     /locus_tag="MSC_0312"
                     /codon_start=1
```

```
KEYWORDS     HTG.
SOURCE       Mus musculus (house mouse)
  ORGANISM   Mus musculus
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE    1 (bases 1 to 196150)
  AUTHORS    Wilson,R.K.
  TITLE      The sequence of Mus musculus clone
  JOURNAL    Unpublished
REFERENCE    2 (bases 1 to 196150)
  AUTHORS    Wilson,R.K.
  TITLE      Direct Submission
  JOURNAL    Submitted (07-NOV-2003) Genome Sequencing Center, 4444 Forest Park
             Parkway, St. Louis, MO 63108, USA
REFERENCE    3 (bases 1 to 196150)
  AUTHORS    Wilson,R.K.
  TITLE      Direct Submission
  JOURNAL    Submitted (09-JUN-2004) Genome Sequencing Center, 4444 Forest Park
             Parkway, St. Louis, MO 63108, USA
REFERENCE    4 (bases 1 to 196150)
  AUTHORS    Wilson,R.K.
  TITLE      Direct Submission
  JOURNAL    Submitted (29-JUL-2004) Genome Sequencing Center, 4444 Forest Park
             Parkway, St. Louis, MO 63108, USA
COMMENT      On Jul 29, 2004 this sequence version replaced gi:48475355.
             --------------- Genome Center ---------------
             Center: Washington University Genome Sequencing Center
             Center code: WUGSC
             Web site:http://genome.wustl.edu
             Contact: submissions@watson.wustl.edu
             -------- Project Information --------
             Center project name: M_BB0109P17
             ---------------------------------------------
FEATURES          Location/Qualifiers
     source       1..196150
                  /organism="Mus musculus"
                  /mol_type="genomic DNA"
                  /db_xref="taxon:10090"
                  /chromosome="13"
                  /clone="RP24-109P17"
```

```
gene         6705..7016
                  /locus_tag="MSC_0313"
CDS          6705..7016
                  /locus_tag="MSC_0313"
                  /codon_start=1
                  /evidence=not_experimental
                  /product="Hypothetical transmembrane protein"
                  /protein_id="CAE76954.1"
                  /db_xref="GI:42492358"
                  /translation="MIFSSCFYSDWFKISLNCVFLMLLVEIHIAMGNENKTIIKQANK
                  DHHCTDIGIEILLITSSLFNNIIGRKILLIVKIITNILKIFVIPKTYQIVCLYDPYFL
                  T"
gene         complement(6967..8808)
                  /locus_tag="MSC_0314"
CDS          complement(6967..8808)
                  /locus_tag="MSC_0314"
                  /codon_start=1
                  /evidence=not_experimental
                  /product="hypothetical transmembrane protein"
                  /protein_id="CAE76955.1"
                  /db_xref="GI:42492359"
                  /translation="MSPYSNNLKTNYQILGLDKTONSFKIDDQLKNKVFLSNSDIQN
                  IYQIINSPYQKQGDDLISSSTNQILYSRKTNTSNVPTILNQTLNNHFLNNQNILDNIS
                  SYNSQLMYKTKSNDFKILPKQAWLYDDTDYLNTKYVTKHAKWMDQPIQSINDQNNYSS
                  YGYKIVENNNKKGHYYLNPYNLDVNKPTQRQVIDLWSDKNENNKDLSIEKQVDNIVENS
                  PIFGDFVINKNGEIIKSFLRPYYQLRNLKLFIPISDEISWEDFAKYALGWGNNQEPND
                  DWKANLEKQDEHTREYIQPALKKLKIDQVPLSVKNAWDSVLTNKNITDYIEVRPYDFS
                  IEQERRIDRRHIYFQFTPGYINIKGVLKSAPTTSLDNLIINGSALFYRRTLGKRKNIA
                  PILKLENNKVNYVNXDLKIKLTNIGVSDIYGKSYAIIDSDLANMIYGYDISRSINYDY
                  KPFDFTSKIIKRNQLFNTYNTTNWQKQNIIDPWKNTYLKSKNVYDYSPHYYNTIFSNS
```

```
Query Match        2.8%; Score 54.8; DB 1; Length 346939;
Best Local Similarity 50.0%; Pred. No. 0.076; Indels 0; Gaps 0;
Matches 137; Conservative 0; Mismatches 137; Indels 0; Gaps 0;

Qy   167 TGGGTGTTCAATTTGTTGATGCTCTCATCAAAAAAGGTCAAATGGCAAAGGAG 226
         || ||| |  ||| |||  ||| |||  | ||||||||| |||| || ||||||
Db 160798 TTGTTGAAAACATTATTGATGAAATTATTAAAACTTCAGAATTCCAAAAAACAA 160857

Qy   227 CATTTAGAGACTCAATTAGAAGTTCTAGAGAAGTACATCCTGATCAATTGATAAGTACA 286
         |||||| ||| || |  |||||||| |||  | || || || |||| |||  ||  |||
Db 160858 TTATTGAAAGACAAATAAAAGATGTTAAAAAGAAAGAATTGAAGATCAATTAGTCAACAAA 160917

Qy   287 AAAAGCTAAAAGTTGATGATTTGGCAGCTGATGCTGATGCAGTTATGCAACAGCGCGAGATGGCAA 346
         |||||  || | |||  ||| |||| |||  | |  ||  |  ||  || ||  |  ||  |||
Db 160918 AAATTACATTAGATAAATATAAAGAAATTACTACTCAAGAGAAATTGATGAAG 160977

Qy   347 AATTACAGCCTAAATCAGGAAATGCATTTATCGATATGTTGAATGGAAATGCCAA 406
         ||| ||  | |||||  ||| |||| || | || ||  | ||  ||  |  ||  |
Db 160978 AACTAAAAAGCGATGCTATTCATAGAATTAAGTCATTTTTAGTTGCAAGTGAAATTAAAA 161037

Qy   407 TTGGTAGCAGTATTCGTTGGATTAGAAGATGCTAT 440
         ||| ||  | |||| | ||| ||||||||||||||
Db 161038 ATAAAGAAAATATTGTTGCAAGTGAAGAAGCTTAT 161071
```

```
RESULT 11
AC147250
LOCUS     AC147250     196150 bp     DNA     linear     ROD 29-JUL-2004
DEFINITION Mus musculus chromosome 13 clone RP24-109P17, complete sequence.
ACCESSION AC147250
VERSION   AC147250.4  GI:50811801
```

```
ORIGIN
Query Match        2.8%; Score 54.6; DB 10; Length 196150;
Best Local Similarity 50.6%; Pred. No. 0.087; Indels 0; Gaps 0;
Matches 132; Conservative 0; Mismatches 129; Indels 0; Gaps 0;

Qy   620 AACAAGCTCCTTGCTCAAAGAGCAATGTTAGGTAAAAATGCCCCAGTTGCCGGTGGAAGAG 679
         ||| |||| | | ||||||||| |||| ||||  |||||| |   ||  || ||  ||||||
Db 56919 AGAAGATGATGAAGAAGAGGAGGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAACTTC 56978

Qy   680 GTGAAGAACAACCGATGATGATGAATCGAGTGATGACCAAAGAATGCAACAAAGAAGAACTTC 739
         || ||||| ||  ||| | |||||  | |||   |||||| | |||||  | ||||| |||
Db 56979 GAAGAAGAAGAAGGAGGAGGAGGAGGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAG 57038

Qy   740 AAGAGGAGATGAAGATGATGATGATCTTGAGGACGACGATGTACCCAGAAGGAAGAAGTT 799
         |||| |  ||| |||| || |||| |  || || ||  | ||  ||||| | || ||  ||
Db 57039 AAGAAGAAGAAGAAGAAGAAGAAGAAGAAGCATCAGAGAAGAAGAGAGAAGGAGGAGGAG 57098

Qy   800 CCGATGGAGAACCACAAAGTGAAGCAGCATCAGAGAAGAATTAGCCAGAGATTGA 859
         || |||||| |  || |  |  ||||||  | |||||||| |  ||||| |||| |
Db 57099 GAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAAGAAGGAGGAGGAGGAGGAAGGAA 57158

Qy   860 AAAGTAGTCCTAGATTAAAAG 880
         |||||||  ||||| |||||
Db 57159 GGAGAAGGAGGAAGAAGAGGAG 57179
```

```
RESULT 12
AC123071/c
LOCUS     AC123071     252384 bp     DNA     linear     ROD 10-JUL-2004
DEFINITION Mus musculus BAC clone RP23-70P8 from chromosome 13, complete
          sequence.
```

```
ACCESSION   AC123071
VERSION     AC123071.3  GI:49258297
KEYWORDS    HTG.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 252384)
  AUTHORS   Tomlinson,C., Bielicki,L. and Haakenson,W.
  TITLE     The sequence of Mus musculus BAC clone RP23-70P8
  JOURNAL   Unpublished (2001)
REFERENCE   2  (bases 1 to 252384)
  AUTHORS   McPherson,J.D. and Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-MAY-2002) Genome Sequencing Center, 4444 Forest Park
            Parkway, St. Louis, MO 63108, USA
REFERENCE   3  (bases 1 to 252384)
  AUTHORS   Wilson,R.K.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-JUN-2004) Genome Sequencing Center, 4444 Forest Park
            Parkway, St. Louis, MO 63108, USA
REFERENCE   4  (bases 1 to 252384)
  AUTHORS   Wilson,R.K.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-JUN-2004) Genome Sequencing Center, 4444 Forest Park
            Parkway, St. Louis, MO 63108, USA
REFERENCE   5  (bases 1 to 252384)
  AUTHORS   Wilson,R.K.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-JUL-2004) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT     On Jun 26, 2004 this sequence version replaced gi:48475394.
            -------------- Genome Center
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site: http://genome.wustl.edu
            Contact: submissions@watson.wustl.edu
            -------------- Summary Statistics
            Center project name: M_BA0070P08
            --------------

            NOTICE:

            This sequence was finished as follows unless otherwise noted:
            all regions were double stranded, sequenced with an alternate
            chemistry, or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by sequence
            from more than one subclone; and the assembly was confirmed by
            restriction digest.

            MAPPING INFORMATION:
            Mapping information for this clone was provided by Dr. Wes Warren,
            Department of Genetics, Washington University, St. Louis MO. For
            additional information about the map position of this sequence, see
            http://genome.wustl.edu

            SOURCE INFORMATION:
            The RPCI-23 BAC Library has been constructed by Kazutoyo Osegawa
            and Minako Tateno in the laboratory of Pieter de Jong
            (http://www.chori.org) from female C57BL/6J mouse kidney and/or
            brain genomic DNA. The clone and detailed information can be
            obtained from Research Genetics, Inc. (http://www.resgen.com) or
            Pieter de Jong and coworkers at http://www.chori.org

            NEIGHBORING SEQUENCE INFORMATION:
            This sequence is the entire insert of the clone.
FEATURES             Location/Qualifiers
     source          1..252384
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
                     /chromosome="13"
                     /map="13"
                     /clone="RP23-70P8"
                     /clone_lib="RPCI-23"
     repeat_region   428..769
                     /rpt_family="ERVK"
     repeat_region   2931..3007
                     /rpt_family="ERVK"
     repeat_region   3024..3307
                     /rpt_family="ERVK"
     repeat_region   3299..4073
                     /rpt_family="ERVK"
     repeat_region   4076..4417
                     /rpt_family="ERVK"
     repeat_region   8061..8455
                     /rpt_family="MaLR"
     repeat_region   10941..11107
                     /rpt_family="B2"
     repeat_region   11852..12148
                     /rpt_family="L1"
     repeat_region   13139..13243
                     /rpt_family="MIR"
     repeat_region   13324..13373
                     /rpt_family="ERV1"
     repeat_region   13351..13423
                     /rpt_family="ERV1"
     repeat_region   13519..13591
                     /rpt_family="B4"
     repeat_region   13592..13649
                     /rpt_family="ID"
     repeat_region   13657..13792
                     /rpt_family="Alu"
     repeat_region   13697..13815
                     /rpt_family="MaLR"
     repeat_region   13793..13839
                     /rpt_family="B4"
     repeat_region   14880..15115
                     /rpt_family="MaLR"
     repeat_region   15122..15310
                     /rpt_family="MER1_type"
     repeat_region   15319..15454
                     /rpt_family="Alu"
     repeat_region   15559..15665
                     /rpt_family="MaLR"
     repeat_region   15886..15945
                     /rpt_family="ERV1"
     repeat_region   15959..16151
                     /rpt_family="B2"
     repeat_region   16800..16902
                     /rpt_family="Alu"
     repeat_region   18719..18859
                     /rpt_family="Alu"
     repeat_region   18945..19413
                     /rpt_family="L1"
     repeat_region   20321..20427
                     /rpt_family="Alu"
     repeat_region   20854..21193
                     /rpt_family="L1"
     repeat_region   21194..21557
                     /rpt_family="MaLR"
     repeat_region   21561..21738
                     /rpt_family="L1"
     repeat_region   22261..22863
                     /rpt_family="RMER6A"
     repeat_region   23612..23752
                     /rpt_family="MIR"
     repeat_region   23808..23935
                     /rpt_family="L2"
     repeat_region   25086..25446
                     /rpt_family="L1"
     repeat_region   25481..25539
                     /rpt_family="ERV1"
     repeat_region   25663..25773
```

```
repeat_region     /rpt_family="MER1_type"
                  25936..26145
repeat_region     /rpt_family="B2"
                  26169..26278
repeat_region     /rpt_family="MER1_type"
                  26546..26700
repeat_region     /rpt_family="B4"
                  27036..27192
repeat_region     /rpt_family="L1"
                  30095..30238
repeat_region     /rpt_family="B4"
                  30330..30464
repeat_region     /rpt_family="Alu"
                  31790..31844
repeat_region     /rpt_family="ERV1"
                  32306..32657
repeat_region     /rpt_family="MaLR"
                  36327..37036
repeat_region     /rpt_family="L1"
                  37065..37371
repeat_region     /rpt_family="MaLR"
                  37443..37719
repeat_region     /rpt_family="MaLR"
                  38157..38662
repeat_region     /rpt_family="MER1_type"
                  40294..40489
repeat_region     /rpt_family="MaLR"
                  40582..40833
repeat_region     /rpt_family="L1"
                  41213..41424
repeat_region     /rpt_family="B2"
                  42174..42233
repeat_region     /rpt_family="MaLR"
                  44135..44500
repeat_region     /rpt_family="MaLR"
                  44502..44814
repeat_region     /rpt_family="ERVK"
                  44819..44867
repeat_region     /rpt_family="MaLR"
                  44963..45023
repeat_region     /rpt_family="ID"
                  47662..47841
repeat_region     /rpt_family="B4"
                  49354..49468
repeat_region     /rpt_family="Alu"
                  49839..50141
repeat_region     /rpt_family="L1"
                  50983..51188
repeat_region     /rpt_family="L1"
                  52932..53027
repeat_region     /rpt_family="B2"
```

```
Db        23026 GGAGAAGAGAAGAAGAGAGAG 23006
                      ||||| ||| ||| ||

RESULT 13
AC120258/c
LOCUS       AC120258      220873 bp    DNA     linear   HTG 12-OCT-2002
DEFINITION  Rattus norvegicus clone CH230-246N20, WORKING DRAFT SEQUENCE.
ACCESSION   AC120258
VERSION     AC120258.3  GI:23908222
KEYWORDS    HTG; HTGS_PHASE2; HTGS DRAFT; HTGS_FULLTOP.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 220873)
AUTHORS     Muzny,D.Marie, Metzker,M.Lee, Abramzon,S., Adams,C., Alder,J.,
            ... [author list] ...
TITLE       Direct Submission
JOURNAL     Unpublished
REFERENCE   2  (bases 1 to 220873)
AUTHORS     Worley,K.C.
TITLE       Direct Submission
JOURNAL     Submitted (05-MAY-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
```

```
Query Match      2.8%;  Score 54.6; DB 10; Length 252384;
Best Local Similarity   50.6%;  Pred. No. 0.086;
Matches  132;  Conservative    0;  Mismatches 129;  Indels    0;  Gaps    0;

Qy    620 AACAAGCTCTTGCTCAAAGAGCAATGTTAGGTAAAAATGCCCCAGTTGCCCGGTGGAAGAG 679
                  ||||| || ||||||||| ||||||| |||||
Db  23266 AAGAAGATGATGAAGAAGAGGAGGAGGAGGAAAAAAGAGGAGGAGGAAGAAGAA 23207

Qy    680 GTGAAGAACAACGGATGATGATGAATCGAGTCGACCAAAGAATGCAACAAAGAGAACTTC 739
                  |||||||| || ||| || ||| || |||||||||
Db  23206 GAAGAAGAGGAGGAGGAGGAGGAGGAAAAGAGGAGGAGGAAGAAGAAGAAG 23147

Qy    740 AAGAGGAAGAGATGAAGATGATGATGATCTTGAGGACGAGGATGTACCCAGAAGAAGAAGTT 799
                  ||||| ||||||| |||| ||||| ||||||||
Db  23146 AAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAGGAGGAGGAGGAGGACGAGCGAG 23087

Qy    800 CGGATGAGAACCACAAAGTGAAGCAGCAGCATCAGAGAAGAGATTTAGCCAGGAGATTGA 859
                  |||||||| || ||||||||| || |||||||||
Db  23086 GAGGAGGAGGAGGAGGAGGAGGAGGAGGAAGAGGAGGAGGAGGAGGAAGAGGAA 23027

Qy    860 AAAGTAGTCCTAGATTAAAAG 880
```

REFERENCE   3 (bases 1 to 220873)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE   Direct Submission
  JOURNAL   Submitted (12-OCT-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT   On Oct 12, 2002 this sequence version replaced gi:21747305.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
            by sized gaps filled with Ns to the estimated size. The sequence
            may extend beyond the ends of the clone and there may be sequence
            contigs within a contig-scaffold that consist entirely of whole
            genome shotgun sequence reads. Both end sequences and whole genome
            shotgun sequence only contigs will be indicated in the feature
            table.
            ----------- Genome Center
            Center: Baylor College of Medicine
            Center code: BCM
            Web site: http://www.hgsc.bcm.tmc.edu/
            Contact: hgsc-help@bcm.tmc.edu
            ----------- Project Information
            Center project name: GVWN
            Center clone name: CH230-246N20
            ----------- Summary Statistics
            Assembly program: Phrap; version 0.990329
            Consensus quality: 197817 bases at least Q40
            Consensus quality: 201230 bases at least Q30
            Consensus quality: 202948 bases at least Q20
            Estimated insert size: 203330; sum-of-contigs estimation
            Quality coverage: 7x in Q20 bases; sum-of-contigs estimation
* NOTE: Estimated insert size may differ from sequence length
*   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
*   consists of 1 contigs. Gaps between the contigs
*   are represented as runs of N. The order of the pieces
*   is believed to be correct as given, however the sizes
*   of the gaps between them are based on estimates that have
*   provided by the submittor.
*   This sequence will be replaced
*   by the finished sequence as soon as it is available and
*   the accession number will be preserved.
*   1 220873: contig of 220873 bp in length.
FEATURES   Location/Qualifiers
     source   1..220873
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-246N20"
     misc_feature   1..1073
                     /note="wgs_contig"
     misc_feature   1..767
                     /note="clone_boundary"
                     clone_end:T7
                     site:MboI
                     end_sequence:RXAAX82TJ"
     misc_feature   1575..2845
                     /note="wgs_contig"
     misc_feature   3347..4452
                     /note="wgs_contig"
     misc_feature   complement(218194..219030)
                     /note="clone_boundary"
                     clone_end:Sp6
                     site:MboI
                     end_sequence:RXAAX82TV"
ORIGIN
Query Match   2.8%;   Score 54;   DB 2;   Length 220873;

Best Local Similarity   51.2%;   Pred. No. 0.12;
Matches 126;   Conservative   0;   Mismatches 120;   Indels   0;   Gaps   0;

Qy   706 CGAGTGGACCAAGAATGCAACAAAGAGAACTTCAAGAGAGAAGAGGAAGAAGATGATGAT 765
          ||||||||||||| ||||||| ||||||||||||||||||||||||||||||||||||
Db 79454 CAAGGTTATGAAACAATTTACAATGGTAACTGGATGATGATGATGATGATGATGATGAT 79395

Qy   766 CTTGAGGACGAGGATGTACCCAGAGAAGAAGAAGTTCGGATGGAGAACCACAAAGTGAAGCA 825
          ||| ||| ||| ||| ||||||||| ||||
Db 79394 GATGATGATGATGATGAGAAGAAGACCCACTGGGGAGATATTGGGAGAAATTCA 79335

Qy   826 GAGCATCAGAGAAGAAGGATTAGCCAGGAGATTGAAAAGTAGTCCTAGATTAAAAGAGCTT 885
          |||| ||||| |||||||||||||||||
Db 79334 GAAAATAATAATGTACATTAAAATCCTGTATTTTATCAAATTGAAACATGCAAGGGGATAT 79275

Qy   886 TTACAGAATGCGGAAGTTCAATCATTGCTCTCTTACCAACGAATGAGGATTCTCCACTG 945
          ||| || ||||||||||||||||||||||
Db 79274 TGACGAATTTCAAGAGTATACCACTTACCAAAGTTAGACCAAAGTGAGATAAACAATTT 79215

Qy   946 AGCAAG 951
          |||
Db 79214 TAAAAG 79209

RESULT 14
AC097214/c
LOCUS   AC097214   225493 bp   DNA   linear   HTG 10-MAY-2003
DEFINITION Rattus norvegicus clone CH230-150M11, *** SEQUENCING IN PROGRESS
            ***
ACCESSION   AC097214
VERSION   AC097214.7 GI:30521196
KEYWORDS   HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE   Rattus norvegicus (Norway rat)
  ORGANISM Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1 (bases 1 to 225493)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abranzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Derano,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Forbes,L., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,C., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Louiseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindattne,M., Mahmoud,H., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S., Parks,K.,
            Nwaokelemeh,O., Okwuonu,G., Olarmpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,

TITLE   Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,R., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 225493)
AUTHORS   Worley,K.C.
TITLE   Direct Submission
JOURNAL   Submitted (12-OCT-2001) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3 (bases 1 to 225493)
AUTHORS   Rat Genome Sequencing Consortium.
TITLE   Direct Submission
JOURNAL   Submitted (10-MAY-2003) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT   On May 10, 2003 this sequence version replaced gi:23617016.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
            by sized gaps filled with Ns to the estimated size. The sequence
            may extend beyond the ends of the clone and there may be sequence
            contig within a contig-scaffold that consist entirely of whole
            genome shotgun sequence reads. Both end sequences and whole genome
            shotgun sequence only contig will be indicated in the feature
            table.
            ----------- Genome Center
            Center: Baylor College of Medicine
            Center code: BCM
            Web site: http://www.hgsc.bcm.tmc.edu
            Contact: hgsc-help@bcm.tmc.edu
            ----------- Project Information
            Center project name: GIOV
            Center clone name: CH230-150M11
            ----------- Summary Statistics
            Assembly program: Atlas 3.0;
            Consensus quality: 205588 bases at least Q40
            Consensus quality: 209426 bases at least Q30
            Consensus quality: 211934 bases at least Q20
            Estimated insert size: 218092; sum-of-contigs estimation
            Quality coverage: 5x in Q20 bases; sum-of-contigs estimation
            ----
          * NOTE: Estimated insert size may differ from sequence length
          *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
          * NOTE: This is a 'working draft' sequence. It currently
          *   consists of 1 contigs. Gaps between the contigs
          *   are represented as runs of N. The order of the pieces
          *   is believed to be correct as given, however the sizes
          *   of the gaps between them are based on estimates that have
          *   provided by the submittor.
          * This sequence will be replaced
          *   by the finished sequence as soon as it is available and
          *   the accession number will be preserved.
          *   1 225493; contig of 225493 bp in length.
FEATURES   Location/Qualifiers
     source   1..225493
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"

     misc_feature   /db_xref="taxon:10116"
                     /clone="CH230-150M11"
                     3680..4488
                     /note="clone boundary
                     clone end:T7
                     site:EcoRI
                     end sequence:RWBAU78TJC"
     misc_feature   89244..90370
                     /note="wgs_contig"
     misc_feature   131230..133655
                     /note="wgs_contig"

ORIGIN

Query Match   2.8%; Score 54; DB 2; Length 225493;
Best Local Similarity 51.2%; Pred. No. 0.12;
Matches 126; Conservative 0; Mismatches 120; Indels 0; Gaps 0;

```
Qy    706 CGAGTGGACCAAAGAATGCAACAAAGAGAGAACTTCAAGAGAGAAGATGAAGATGATGAT 765
          |||  ||||| | |||  |||| | ||||||||| ||||  ||| |||||||||||||||
Db 209175 CAAGTTATGAAACAATTTACATGGTAACTGATGATGATGATGATGATGATGAT 209116

Qy    766 CTTGAGGACGAGGATGTACCCAGAGAAGAAGAAGTTCGGATGGAGAACCACAAAGTGAAGCA 825
          ||| ||  ||||| |||  |||||||  |||||
Db 209115 GATGATGATGATGATGATGAAGAGAGACCCCTGGGGGAGATATTGGGAGAAATTCA 209056

Qy    826 GAGCATCAGAGAGAGATTTAGCCAGGAGATTGAAAAGTAGTCCTAGATTAAAAGAGCTT 885
          ||| | |||  ||| ||||
Db 209055 GAAATAATAAATGTACATTAAAATCCTGTATTTTATCAAATTGAAACATCCAAGGGATAT 208996

Qy    886 TTACAGATGCGGAAGTTCAATCATTGCTCCTCTTACCAACGAATGAGGGATTCTCCACTG 945
          || |||||  || ||
Db 208995 TGACGAATTTCAAGAGGTATACCACTTACCAAAGTTAGACCAAAGTGAGATAAACAATTT 208936

Qy    946 AGCAAG 951
          ||
Db 208935 TAAAAG 208930
```

RESULT 15
AC113937
LOCUS   AC113937   225983 bp   DNA   linear   HTG 10-MAR-2004
DEFINITION Mus musculus chromosome 6 clone RP23-21F2 map 6, *** SEQUENCING IN
            PROGRESS ***, 9 ordered pieces.
ACCESSION   AC113937
VERSION   AC113937.6 GI:45332531
KEYWORDS   HTG; HTGS_PHASE2; HTGS_FULLTOP; HTGS_ACTIVEFIN.
SOURCE   Mus musculus (house mouse)
ORGANISM   Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 225983)
AUTHORS   Birren,B., Nusbaum,C. and Lander,E.
TITLE   Mus musculus chromosome 6, clone RP23-21F2
JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 225983)
AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
            Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
            Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R.,
            Landers,T., Lehoczky,J., Levine,R., Liu,G., MacLean,C.,
            Macdonald,P., Major,J., Marquis,N., Matthews,C., McCarthy,M.,
            McEwan,P., McKernan,K., Meldrim,J., Meneus,L., Mihova,T.,
            Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C.,
            Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J.,
            Peterson,K., Phunkhang,P., Pierre,L., Pollara,V., Raymond,C.,
            Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J.,
            Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S.,
            Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,

```
                                                                    /mol_type="genomic DNA"
                                                                    /db_xref="taxon:10090"
                                                                    /chromosome="6"
                                                                    /map="6"
                                                                    /clone="RP23-21F2"
                                                                    /clone_lib="RPCI-23 Female Mouse BAC"

ORIGIN

         Query Match           2.8%;   Score 54;   DB 2;   Length 225983;
         Best Local Similarity  38.1%;  Pred. No. 0.12;
         Matches 182;  Conservative  0;  Mismatches  295;  Indels  1;  Gaps  1;

Qy    620 AACAAGCTCTTGCTCAAAGAGCAATGTTAGGTAAAAATGCCCAGTTGCCGTGGAAGAG 679
          |||   |||  |||   |||  |  |  | ||  |  ||| |  | |   |||  |
Db  60237 AAGAAGAAGAACAAGAAGAAGAAGAACAAGAAGAAGAAGAAGAAGAACAAAG 60296

Qy    680 GTGAAGAACAACGGATGATGATGAATCGAGTGGACCAAAGAAATGCAACAAAGAGACTTC 739
          |||   |||  |||   |||  | || |  | ||  | ||   ||||
Db  60297 AAGAACAAAAGAAAAAAACAACAAGAAGAACAACAAGAAGAACAAAG 60356

Qy    740 AAGAGGAAGATGAAGATGATGATGATCTTGAGGACGCAGGATGTACCCAGAAGGAAGTT 799
          |||  |||  ||||| |  |   |||  |  | || |  | ||  ||||||
Db  60357 AAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAGATTTAGCCAGGAGATTGA 60415

Qy    800 CGGATGAGAACCACAAAGTGAAGCAGAGCATCAGAGAAGGACATCNNNNNNNNNNNNNNNNN 859
          |||   |  ||||
Db  60416 GAAGAAGAACGTGTTATGAAAAAGAAGAAT CNNNNNNNNNNNNNNNNNNNNNNNNNNNN 60475

Qy    860 AAAGTAGTCCTAGATTAAAAGAGAGCTTTTACAGAATGCGGAGTTCAATCATTGCTCTCTT 919
          NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
Db  60476 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 60535

Qy    920 ACCAACGAATGAGGGATTCTCCACTGAGCAAGCGAGGGCCTTTGGCTATGACGATGCAGG 979
          | ||||  | || |||   || ||   |||  | ||   || |  |  ||  |  ||
Db  60536 NNNNNNNNNNNNGAAAAGAAGAAGAAGAAGAAGAAGACGAAGAAGACGAAGAGAGA 60595

Qy    980 ATGAAAGTGCATTCCGGCAATGGAGGCTCGTGCAAAATCTAGATCAAAAATCTCAACTTG 1039
          |||   |||  | |||  |||  ||   |||  |||
Db  60596 AGGAAGAAGAAGAAGAAGACGAAGAAGAACGAAGAAGAAGAGAAGAAGAAGAAG 60655

Qy   1040 TGCTCGGTCTCCATGGTTTTGGAGAGTTCTGATGATGAAGACGATGATGAAGACGAAGAAAA 1097
          ||||||   ||   |||   |||  ||||| |  || ||  ||||||
Db  60656 AAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAACGAAGAAGACGAAGA 60713


         Search completed: July 15, 2005, 13:09:15
         Job time : 8512 secs
```

Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
Topham,K., Travis,M., Travis,N., Trigilio,J., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

TITLE   Direct Submission
JOURNAL   Submitted (05-MAR-2002) Whitehead Institute/MIT Center for Genome
           Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE 3 (bases 1 to 225983)
AUTHORS   Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,
           Anderson,M., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T.,
           Boguslavkiy,L., Boukhgalter,B., Camarata,J., Chang,J., Choepel,Y.,
           Collymore,A., Cook,A., Cooke,P., Corum,B., DeArellano,K.,
           Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S.,
           Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S.,
           Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B.,
           Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
           Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
           Lindblad-Toh,K., Liu,X., Lui,A., Mabbitt,R., MacLean,C.,
           Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M.,
           Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J.,
           Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P.,
           O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
           Rachupka,A., Ramasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P.,
           Roman,J., Schauer,S., Schupback,R., Seaman,S., Severy,P., Smith,C.,
           Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M.,
           Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M.,
           Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X.,
           Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
TITLE   Direct Submission
JOURNAL   Submitted (10-MAR-2004) Whitehead Institute/MIT Center for Genome
           Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT   On Mar 10, 2004 this sequence version replaced gi:44886384.
           All repeats were identified using RepeatMasker.
           Smit,A.F.A. & Green, P. (1996-1997)
           http://ftp.genome.washington.edu/RM/RepeatMasker.html
           --------------------- Genome Center
             Center: Whitehead Institute/ MIT Center for Genome Research
             Center code: WIBR
             Web site: http://www-seq.wi.mit.edu
             Contact: sequence_submissions@genome.wi.mit.edu
           --------------------- Project Information
             Center project name: L23232
             Center clone name: 21_F_2
           ----
         * NOTE: This is a 'working draft' sequence. It currently
         * consists of 9 contigs. Gaps between the contigs
         * are represented as runs of N. The order of the pieces
         * is believed to be correct as given, however the sizes
         * of the gaps between them are based on estimates that have
         * provided by the submittor.
         * This sequence will be replaced
         * by the finished sequence as soon as it is available and
         * the accession number will be preserved.
         *   1   12243: contig of 12243 bp in length
         *   12244   12343: gap of 100 bp in length
         *   12344   44783: contig of 32440 bp in length
         *   44784   44883: gap of 100 bp in length
         *   44884   60448: contig of 15565 bp in length
         *   60449   60548: gap of 100 bp in length
         *   60549   82671: contig of 22123 bp in length
         *   82672   82771: gap of 100 bp in length
         *   82772   127514: contig of 44743 bp in length
         *   127515   127614: gap of 100 bp in length
         *   127615   162894: contig of 35280 bp in length
         *   162895   162994: gap of 100 bp in length
         *   162995   185023: contig of 22029 bp in length
         *   185024   185123: gap of 100 bp in length
         *   185124   192636: contig of 7513 bp in length
         *   192637   192736: gap of 100 bp in length
         *   192737   225983: contig of 33247 bp in length.
FEATURES   Location/Qualifiers
     source   1..225983
                     /organism="Mus musculus"

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 15, 2005, 10:39:15 ; Search time 1071 Seconds
(without alignments)
10678.749 Million cell updates/sec

Title: US-10-736-868-1
Perfect score: 1932
Sequence: 1 atgattttatttatttt........tgcctgcactgggattgtag 1932

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 439206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :   N_Geneseq_16Dec04:*
1: geneseqn1980s:*
2: geneseqn1990s:*
3: geneseqn2000s:*
4: geneseqn2001as:*
5: geneseqn2001bs:*
6: geneseqn2002as:*
7: geneseqn2002bs:*
8: geneseqn2003as:*
9: geneseqn2003bs:*
10: geneseqn2003cs:*
11: geneseqn2003ds:*
12: geneseqn2004as:*
13: geneseqn2004bs:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		57.2	3.0	2000	8	ADA71938	Rice gene
2		51.6	2.7	3653	2	AAX20280	Borrelia
3		51.2	2.7	3399	2	AAT05868	Chicken 1
4	c	50.4	2.6	193853	11	ACN44956	Mouse gen
5	c	49.4	2.6	110000	2	AAV21209_01	Continuation (2 of
6		48.2	2.5	77287	9	AAD58261	Murine tu
7		48.2	2.5	167739	9	AAD58258	Murine tu
8		47.8	2.5	13407	13	ABD33029	Mouse can
9	c	47.8	2.5	180385	10	ADL13931	Osteoarth
10	c	47.2	2.4	3534	8	ACA28380	Prokaryot
11		47	2.4	1686	2	AAQ87587	DNA encod
12	c	46.2	2.4	2427	5	AAB88862	DNA encod
13	c	46.2	2.4	2427	5	AAS69554	DNA encod
14		46	2.4	423	5	AAS90685	DNA encod
15	c	45.8	2.4	17848	4	AAB45323	Chemical
16	c	45.8	2.4	17848	6	ABK39976	Human che
17	c	45.8	2.4	17848	6	ABK28164	DNA trans
18		45.8	2.4	127767	13	ABD32657	Mouse can
19	c	45.4	2.3	272	4	AAI21218	Probe #11
20	c	45.4	2.3	272	4	ABA66298	Human foe
21	c	45.4	2.3	272	4	AAI46489	Probe #15
22	c	45.4	2.3	272	4	ABA48402	Human bre
23	c	45.4	2.3	272	4	ABA33364	Probe #11
24	c	45.4	2.3	272	4	AAK40459	Human bon
25	c	45.4	2.3	272	4	AAK14716	Human bra
26	c	45.4	2.3	272	4	ABS40022	Human liv
27	c	45.4	2.3	272	5	AAI06928	Probe #69
28	c	45.4	2.3	798	6	ABS14451	Human gen
29	c	45.4	2.3	1407	5	AAS90700	DNA encod
30	c	45.4	2.3		5	AAS90722	DNA encod
31	c	45.4	2.3	153170	12	ADQ17382	Human sof
32	c	45.2	2.3	98638	12	ADQ97919	Mouse can
33	c	44.8	2.3	543	12	ACH70417	Human gen
34	c	44.6	2.3	61876	11	ACN44732	Mouse gen
35	c	44.4	2.3	32069	8	AAD56078	Mouse Map
36	c	44.4	2.3	32069	9	ADA02440	Mouse Map
37	c	44.4	2.3	32069	10	ADB72179	Mouse Map
38	c	44.4	2.3	224112	13	ABD32600	Mouse can
39	c	44.2	2.3	393	5	AAS93402	DNA encod
40	c	44.2	2.3	1215	5	AAS92075	DNA encod
41	c	44.2	2.3	1218	5	AAS90730	DNA encod
42	c	44.2	2.3	61791	13	ABD33484	Murine ca
43	c	44.2	2.3	110000	13	ABD32594_2	Continuation (3 of
44	c	44	2.3	110000	3	AAF22303	Arabidops
45	c	43.8	2.3	379	5	AAS90909	DNA encod

ALIGNMENTS

RESULT 1
ADA71938
ID   ADA71938 standard; DNA; 2000 BP..
XX
AC   ADA71938;
XX
DT   20-NOV-2003 (first entry)
XX
DE   Rice gene, SEQ ID 5263.
XX
KW   Plant; bacterial infection; fungal infection; viral infection; rice;
KW   gene; ds.
XX
OS   Oryza sativa.
XX
PN   WO200300898-A1.
XX
PD   03-JAN-2003.
XX
PF
XX
PR   22-JUN-2001; 2001WO-IB001105.
PR   22-JUN-2001; 2001WO-IB001105.
XX
PA   (SYGN) SYNGENTA PARTICIPATIONS AG.
XX
PI   Chang H, Chen W, Cooper B, Glazebrook J, Goff SA, Hou Y;
PI   Katagiri F, Quan S, Tao Y, Whitham S, Xie Z, Zhu T, Zou G;
XX
DR   WPI; 2003-175290/17.
XX
PT   Identifying at least one gene involved in plant resistance or response to
PT   pathogenic infection for conferring resistance or tolerance to a plant to
PT   bacterial, fungal or viral infection by determining or detecting plant
PT   gene expression.
XX
PS   Claim 27; SEQ ID NO 5263; 899pp; English.
XX
CC   The present invention relates to a method (M1) for identifying genes
CC   involved in plant resistance or response to pathogenic infection. M1
CC   comprises identifying a gene whose expression is significantly altered in
CC   the incompatible interaction of plant gene expression relative to
CC   expression of the gene in an uninfected plant, in a mutant plant that
CC   does not express a gene associated with response to pathogenic infection,

CC   or in a corresponding incompatible or compatible interaction. (M1) is
CC   useful for conferring resistance to resistance or tolerance to a plant to
CC   bacterial, fungal or viral infection. The present sequence was used to
CC   illustrate the invention.
XX
SQ   Sequence 2000 BP; 336 A; 265 C; 284 G; 363 T; 0 U; 752 Other;

Query Match   3.0%;   Score 57.2; DB 8; Length 2000;
Best Local Similarity 10.8%; Pred. No. 0.00045;
Matches   74; Conservative 322; Mismatches 280; Indels 12; Gaps 3;

Qy   726 ACAAAGAGAACTTCAAGAGAGAAGATGAAGATGATGAGATCTTGAGGACGAGGATGTACC 785
          :::::::::::: ::::: :|::::||::::|:|:::::::::::::::::::::::::
Db   346 RMMMRSMCRWSKACYYMRWRMWRMTRRRRWAKKSSRTSRRKRKWCMRKRKYKRMRGYSR 405

Qy   786 CAGAAGAAGAAGTTCGGATGGAGAACCACAAAGTGAAGCAGCATCAGAGAAGAGATTT 845
          :::: :::::: :|::::: ::|:::::::::: ::: ::: ::::::::::::::::::
Db   406 MRSCKRARWMKRCRSGRAWMGCRGCMTCRMKSYGMMRWKSWKRMASKYKWMSRMYRWRK 465

Qy   846 AGCCAGGAGATTGAAAAGTAGTCCTAGATTAAAAGAGCTTTTACAGAATGCGGAAGTTCA 905
          ::::: :::::::::::|:|||:::: :: :::::::: ::: ::::|::::: :::::::
Db   466 KKCSRTTMWGKTRGGMMGTMGRCRYKKRSGMKRKCRRRRWGRMYRMWKRYYMSARYTMR 525

Qy   906 ATCATTGCTCTCTTACCAACGAATGAGGGGATTCTCCACTGACGCAAGCGAGGCCTTTGGC 965
          :: :|::::::|||: ::::::|: ::: ::: ::::: ::: :::::|:::::::::::
Db   526 YCARKKYSYSARKARCWYRGKGYYWAGMWMKRYKRMYMYKMMWYKRKYSKCSWYCKMS 585

Qy   966 TATGAACGATGAGGATGAAAGTGCATTCCGCCGATGGAGGCTCGTGCAAAACTAGATCA 1025
          :::: ::::::|||:::::: : ::::::: :::|: ::: ::::::::::::::::::::
Db   586 YYASCMSKARKAGAKMCKRSKMSAWSKSMRSSRKCRKCASKRSSAKRYANMG--GMTSG 642

Qy   1026 AAAATCTCAACTTGTCCTCCGGTCTCCATGGTTTGGAGAGTCTGATGATGAAGACGA 1085
          :: :::|::::: ::::|||::: ::::: ::: ::: :::::::::: :::::::
Db   643 SRMSRWKSYTCYWRKWGSMKSTCTWMYYMSKYTYAKYGSYWRY------RYRAWCMYMW 695

Qy   1086 AGAAGATGAAAATTTGATTGATCCATCTGAAAATTCATTCCGTCGTGCACCACTTCGTCT 1145
          ::::::::::::::: : ::::::: ::::::|::::::::::::::::::::::::::::
Db   696 RWYYRYRSYMTYMAWYTSSTRMAMTGMKYSGRYWTSWYKYCKCSWKYRSMWYYWSWWWA 755

Qy   1146 TCTTCCGGATTCGTTGAGAAATTAAAGTCAAATGAATTAAAGTGCATTGGACAG 1205
          :|:||:::|:||::::::::: ::::::::::: :::::::::::::::|::::::
Db   756 KTWKMWRRYATRMWMWMYRYSMKWYTWCTMWGYWWYWWRTYMKMRYMWYCTKTYWYWSA 815

Qy   1206 AATTAAATATCGAGTTGATGACGTGGAAAAGTATCTTGCTCCAAAGCCGATGGAATTCAA 1265
          :::::::::: ::: :::::::::::::::::: :: ::: ::::: :::::::::::::
Db   816 TYWTGTWAAWWMAKTKMRMGMTGAKTRGRARKARYWKWATWCATKRWMTKGKGAKWAWT 875

Qy   1266 TCCAAAACCTCAGCCTGGCTACTTTGCTCCACGTAAAATCCACACAAGACCACGTAAAAT 1325
          :::::::::: ::: ::::::::::: ::: ::::::|::::: :::::::::::::||||
Db   876 WMAKAWRKYYWSWM--RAWYYYYKTRRTRYRTCWKARWGSWAYWRMWWNKGSAKMWWMWK 933

Qy   1326 GCTTCCATTATTAATTGGATCTGATCCAAAAGTTCAAGAGGAAAATACGAAGACATCCAAG 1385
          :::::::::::::: : :::::::::::::::::::::::: ::::::::::::::::::::
Db   934 GGRWGWTKYWYWYCTTWKMACGRATKYMCCAGWAMYSYSWTRTYWMRTWRWMWASSRTA 993

Qy   1386 TACCGAATGGAAAATTGCAAAAGAATCA 1413
          ::::: ::: :: ::::::::::::::||
Db   994 KRMARMMWKTRAWSKSYARAYWKMAGCA 1021

RESULT 2
AAX20280
ID   AAX20280 standard; DNA; 3653 BP.
XX
AC   AAX20280;
XX
DT   04-MAY-1999 (first entry)
XX
DE   Borrelia burgdorferi polynucleotide sequence #33.
XX
KW   Borrelia burgdorferi; spirochete; bacterium; pathogen; Lyme disease;
KW   epidemic relapsing fever; endemic relapsing fever; Lyme borreliosis;
KW   infection; diagnosis; characterisation; detection; ds.

XX   Borrelia burgdorferi.
OS
XX
PN   WO9858943-A1.
XX
PD   30-DEC-1998.
XX
PF   18-JUN-1998;   98WO-US012764.
XX
PR   20-JUN-1997;   97US-0050359P.
PR   22-JUL-1997;   97US-0053344P.
PR   22-JUL-1997;   97US-0053377P.
PR   03-SEP-1997;   97US-0057483P.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
PA   (MEDI-) MEDIMMUNE INC.
XX
PI   Fraser C,   White OR,   Clayton R,   Dougherty BA,   Lathigra R;
PI   Smith HO;
XX
DR   WPI; 1999-081217/07.
XX
PT   New isolated Borrelia burgdorferi nucleic acids - used to develop
PT   products for the detection, diagnosis, characterisation, prevention and
PT   therapy of infections, particularly Lyme disease.
XX
PS   Claim 1; Page 994-996; 1128pp; English.
XX
CC   AAX20248 to AAX20402 represent polynucleotide sequences isolated from
CC   Borrelia burgdorferi (Bb). Products derived from Bb can be used for the
CC   detection, diagnosis, characterisation, prevention and therapy of Bb
CC   infections, e.g. Lyme disease. They can also be used for the production
CC   of biosynthetic products, e.g. enzymes. Borrelia belongs to a family of
CC   motile, spiral-shaped bacteria called Spirochetes. Spirochetes are
CC   pathogenic in humans and Borrelia causes epidemic and endemic relapsing
CC   fever, and Lyme borreliosis, more commonly known as Lyme disease.
XX
SQ   Sequence 3653 BP; 1587 A; 348 C; 690 G; 1027 T; 0 U; 1 Other;

Query Match   2.7%;   Score 51.6; DB 2; Length 3653;
Best Local Similarity 48.0%; Pred. No. 0.016;
Matches 147; Conservative 0; Mismatches 159; Indels 0; Gaps 0;

Qy   543 TATGATCGCTGGCAAGAACCCCTTAAAATGCCTCAACAAATGAGAAAAGCTCAAGCTGC 602
          ||| ||| || || ||| || |||||| || ||||||||| | |||| ||| |||
Db   1280 TAATAGTGGTGTAATAAATTCGTCACCAGTATTGCCAGAAAATAGTCAAGATAATACACC 1339

Qy   603 TCCATCGTCAGTGTTCCAACAGCTCTTGCTCAAGAAGCAATGTTAGGTAAAAATGCCCC 662
          |||| || | ||||| || || | || |||||||| ||| ||||| ||||| ||||
Db   1340 AATATTAAAAGCAGCGGAACAAAGTGATGGTCAACAAGAGAAGAAAAAGTGAAAAAGTAGA 1399

Qy   663 AGTTGCCGGTGGAGAGGTGAAGAACAACGGATGATGAGTGGAATCGAGTCAGAGAAGAAT 722
          ||| |||| || || |||| ||| || | || ||||||| || |||||| |
Db   1400 AGAATCCGAAGCTAAAGTTGAGGGAAAAAGAAGAATACAGAACGAACGAAA 1459

Qy   723 GCAACAAAGAACTTCAAGAGAGGAAGATGAAGAATGATGATCTTGAGGACGAGGAGGATGT 782
          || || ||||| | | | ||||| |||||| | || | ||||| ||| || || |
Db   1460 CAACAAGAATTAGCTAAACAAGAAGAACAACAACAAAAGGCAAGAAAA 1519

Qy   783 ACCCAGAAGAAGAAGTTCGGATGGAGAACCACAAAGTGAAGCAGCATCAGAGAAGAAGA 842
          ||| ||||||||| || || ||| || || |||| ||| || || || || ||||||||
Db   1520 ACAAAAAAGAAGAAGAACAAAGACAAAAAGGAAGAAGAGCAAGAAAAGC 1579

Qy   843 TTTAGC 848
          ||||||
Db   1580 TAAGGC 1585

RESULT 3
AAT05868
ID   AAT05868 standard; DNA; 3399 BP.
XX
AC   AAT05868;

XX
DT   27-AUG-2003 (revised)
DT   14-AUG-1996 (first entry)
XX
DE   Chicken leucocytozoan DNA encoding immunogenic protein for vaccines.
XX
KW   Chicken leucocytozoan; immunogen; recombinant vaccine; protection;
KW   immunisation; vaccination; ss.
XX
OS   Leucocytozoon sp.
XX
FH   Key   Location/Qualifiers
FT   CDS   1..3399
FT   /*tag= a
FT   misc_feature   1150..3218
FT   /*tag= b
FT   /note= "fragment referred to in the claims, for use as
FT   insert in a recombinant vaccine against chicken
FT   leucocytozoon disease"
XX
PN   JP07284392-A.
XX
PD   31-OCT-1995.
XX
PF   19-APR-1994; 94JP-00080643.
XX
PR   19-APR-1994; 94JP-00080643.
XX
PA   (DOBU-) DOBUTSUYO SEIBUTSUGAKUTEKI SEIZAI KYOKAI.
PA   (KITA) KITASATO KENKYUSHO SH.
XX
DR   WPI; 1996-006311/01.
DR   P-PSDB; AAR97866.
XX
PT   Chicken leucocytozoan immunogenic protein - used in a recombinant vaccine
PT   against chicken leucocytozoon disease.
XX
PS   Claim 6; Page 6-9; 35pp; Japanese.
XX
CC   AAT05868 encodes a chicken leucocytozoan immunogenic protein, this DNA or
CC   a fragment of it can be used in a recombinant vaccine to immunise against
CC   chicken leucocytozoon disease. The DNA is used in a vector and
CC   operatively linked to an expression regulatory sequence as in standard
CC   practice. (Updated on 27-AUG-2003 to correct OS field.)
XX
SQ   Sequence 3399 BP; 1577 A; 508 C; 798 G; 516 T; 0 U; 0 Other;

Query Match   2.7%; Score 51.2; DB 2; Length 3399;
Best Local Similarity   43.6%; Pred. No. 0.02;
Matches 278; Conservative   0; Mismatches 358; Indels   2; Gaps   1;

Qy   483 GATTGCCAAAGCCGTAATGGACAAATTTCAAACACAAATTCTCCCAGGACTCGTTGCAAA 542
            |||| ||||| || ||| ||||| ||| || || ||| | ||| || |||
Db   1570 GATTTACCAGAAGTAAATGTAGAAGAATCAGATGAAACAGAACATATAACATATGAAATA 1629

Qy   543 TATGATCGCTGGCAAGAACCCCTTTAAAA--TGCCTCAACAAATGAGAAAAGTCAAGCT 600
            |||| || ||| ||||| || |||| ||||| || ||| ||| ||| |||
Db   1630 GATAATGACATACAAGAAGACATGAAAAAGTAACAACATGAAGAAGAAGTAAAAGAGTA 1689

Qy   601 GCTCCATCGTCAGTGTTCCAACAGCTCTTGCTCAAGAGCAATGTTAGGTAAAAATGCC 660
            | ||| ||| || |||| | ||| || ||| || ||| || || || |||
Db   1690 ACACATGAAGAAATAGAAAAGAAGAGCATGAAGAACAACGATGAATGATGCGAGCAAGA 1749

Qy   661 CCAGTTGCCGTCGAAGAGGTGAAGAACAACGATGATGATGAATCGAGTCGACCAAAGA 720
            || || | ||| ||| | ||| | || ||| || | || ||| ||
Db   1750 GAAGTAACCAGTAACATGAAGAAGAAATAGAAAAGAAGAAGCATGAAGAAGTAATACAT 1809

Qy   721 ATGCAACAAAGAGAACTTCAAGAGGAAGATGAAGATGATGATCTTGAGGACGAGGAT 780
            |||| ||| || | || ||| || || || || || || ||| | || ||
Db   1810 AAAGAAGTAACACATGAAGAAGAAATAGAAAAGAAGAGCATCGAAGAGTAATACATGAA 1869

Qy   781 GTACCCAAAGAAGAAGTTCGGATGTGGAGGAGAACCACAAAAGTGAAGCAGCAGAGAGA 840
            | | ||| ||||| || || | ||| |||| || || || || | || ||
Db   1870 GAAGAAAAAGAAGAAGTAACATGAAGAATAGAAAAGAAGATAGAGGAAGTAATA 1929

Qy   841 GATTTAGCCAGGAGATTGAAAAGTAGTCCTAGATTAAAAGAGCTTTTACGAAATGCGGAA 900
            ||| || |||| || || || || | ||| | || ||| || || | || |
Db   1930 CATGAGAAGAAAAAGAAGAAGTAACACATGAAGAATAGAAAAGAAGAGCATGAAGAAGAA 1989

Qy   901 GTTCAATCATTGCTCTTACCAACGAATGAGGGATTCTCCACTGAGCAAGCGAAGGCCT 960
            ||| | || || || | | || |||| || || || || ||| || | || |
Db   1990 GTAATACATGAGGAGAATAGAAAAGAAGTAACACATGAATAGAAAATAGAAGAGAGCAT 2049

Qy   961 TTGGCTATGAACGATGAGGATGAAGTGCATTCCGCGCAATGGGAGCTCGTGCAAAACTA 1020
            || ||||| || ||||| ||| || || || || || ||| || || | ||||
Db   2050 GAAGAAGTAATACATGAAGAAGAAAGAAGAAGTAACACATGAAGAAATAGAAAAAGAA 2109

Qy   1021 GATCAAAAATCTCAACTTGTGCTCGGTCCATGGTTTCCATGAGTGTCGTGATGATGAA 1080
            ||||| || || || | || ||| | || || | || || | ||| ||||
Db   2110 GAGCATGAGAAGAAGTAATACATGAAGAAGAAGTAACACATGAAGAAATAGAAAAAGAA 2169

Qy   1081 GACGAAGAAGATGAAAAATTTGATTGATCCATCTGAAA 1118
            || ||| | |||| ||||| || | | ||||||
Db   2170 AAAGAAGAGCCATGAAGAAGTAATAATCATGAAGAGAAA 2207

RESULT 4
ACN44956/c
ID   ACN44956 standard; DNA, 193853 BP.
XX
AC   ACN44956;
XX
DT   18-NOV-2004 (first entry)
XX
DE   Mouse genomic sequence mCG6056.
XX
KW   Cytostatic; carcinoma; lymphoma; cancer; murine; gene; ss.
XX
OS   Mus musculus.
XX
PN   WO2003073826-A2.
XX
PD   12-SEP-2003.
XX
PF   28-FEB-2003; 2003WO-US006235.
XX
PR   01-MAR-2002; 2002US-00087192.
XX
PA   (SAGR-) SAGRES DISCOVERY.
XX
PI   Morris DW;
XX
DR   WPI; 2003-328604/31.
XX
PT   Recombinant nucleic acid useful for diagnosis and treatment of carcinoma
PT   comprises a nucleotide sequence.
XX
PS   Claim 1; SEQ ID NO 1663; 0pp; English.
XX
CC   The present invention relates to novel DNA and protein sequences which
CC   are associated with carcinomas. The sequences are useful for: (i) for
CC   screening drug candidates; (ii) for screening of bioactive agent capable
CC   of binding to Carcinoma Associated Protein (CAP); (iii) for screening of
CC   a bioactive agent capable of modulating the activity of CAP; (iv) for
CC   evaluating the effect of a candidate carcinoma drug; (v) for diagnosing
CC   carcinoma; (vi) for inhibiting the activity of CAP; (vii) for treating
CC   carcinoma; (viii) for neutralizing the effect of CAP; (ix) as a biochip;
CC   (x) for diagnosing carcinoma or a propensity to carcinoma; and (xi) for
CC   determining Carcinoma Associated (CA) gene copy number. In addition, the
CC   CA genes are useful as DNA vaccines and the CAP are useful as markers of
CC   carcinoma including lymphoma. The present sequence is one such CA coding
CC   sequence. Note: This patent is an equivalent to basic patent
CC   US20021825861, for which no sequence data was published
XX
SQ   Sequence 193853 BP; 44192 A; 37310 C; 40327 G; 53970 T; 0 U; 18054 Other;

Db   80019 TTGCCTGGGCTGAAAAAGGGAAATATTATATAGAGAAGGCAAACTTAAAAAATCATTAG 79960
            |||| ||||||| ||| |||| | || |||||

Qy   1340 TTGGATCTGATCCAAAAGTTCAAGACGAATCGAGTGGACCAAAGAATGCAACAAAGA 1399
            ||||| || |||||||| || |||||| |||| || || ||

Db   79959 AATGTTTTGAT--AATGCTTTAAAAAATAAATCCCAAAGATTGCCAGTCCTTATTGTACAA 79902

Qy   1400 TTGCAAAGAATCAAGAGTTTGACAAATTTGAAGAATAATCCAAGTTCTTGCTGCATTGT 1459
            |||| || |||||||| ||| || ||| || |||| ||

Db   79901 AGGAGAAATATATTTAAACTTGGAGATATGGAGAGCATTAAAAAATGTCTAAAAAAAGT 79842

Qy   1460 TCATGGATGATAAATTAGAGA 1480
            ||||| ||||| ||||||

Db   79841 TTTTGAAAGAAATAATAAAGA 79821

RESULT 6
AAD58261
ID   AAD58261 standard; DNA; 77287 BP.
XX
AC   AAD58261;
XX
DT   20-NOV-2003 (first entry)
XX
DE   Murine tumour suppressor gene, Lmt intron 3 DNA.
XX
KW   Tumour suppressor gene; Lmt; cancer; therapy; cytostatic; murine; ds.
XX
OS   Mus sp.
XX
PN   WO2003066869-A1.
XX
PD   14-AUG-2003.
XX
PF   07-FEB-2003; 2003WO-AU000126.
XX
PR   07-FEB-2002; 2002AU-00000371.
XX
PA   (HALL-) HALL INST MEDICAL RES WALTER & ELIZA.
XX
PI   Cook WD, Mccaw BJ;
XX
DR   WPI; 2003-646311/61.
XX
PT   New nucleic acid molecule, useful for screening a subject for the
PT   presence of an aberration in a gene encoding an LMT.
XX
PS   Claim 10; Page 196-218; 373pp; English.
XX
CC   The invention relates to novel tumour suppressor gene, referred to as
CC   Lmt. The invention also relates to the field of cancer therapy and cancer
CC   diagnostics. The nucleic acid molecule is useful for screening a subject
CC   for the presence of an aberration in a gene encoding an LMT. The present
CC   sequence is murine Lmt intron 3 DNA
XX
SQ   Sequence 77287 BP; 20906 A; 15437 C; 15922 G; 20915 T; 0 U; 4107 Other;

Query Match   2.5%; Score 48.2; DB 9; Length 77287;
Best Local Similarity 55.8%; Pred. No. 0.45;
Matches   92; Conservative 0; Mismatches 73; Indels 0; Gaps 0;

Qy   717 AAGAATGCAACAAAGAAGAACTTCAAGAGGAAGATGAAGAAGATGAATCTTGAGACGA 776
            || | |||||||||||||||||||||||||||||||||||||||| || ||||||

Db   41785 AAGGAGGAAGGAGAGGAAGAAGGAAGAAGGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGA 41844

Qy   777 GGATGTACCCAGAAGAGTTCGGATGAGAACCACAAAGTGAAGCAGAGCATCAGAG 836
            || |||||||||||||| || |||| |||| ||||||

Db   41845 AGAAGAAGAAGAAGAAGAAGAAGAAGAACAAAGAAGGAGGAAGGAAGAAGGAAG 41904

Qy   837 AAGAGATTTAGCCAGGAGATTGAAAAGTAGTCCTAGATTAAAAGA 881
                ||||| |||||| ||| |||||

Db   41905 GAAGGAGGAAGAAGAAGAAGAGGAGAAGAGAAAAGAAGAAAAGAAGA 41949

Query Match   2.6%; Score 50.4; DB 11; Length 193853;
Best Local Similarity 28.9%; Pred. No. 0.18;
Matches 123; Conservative 0; Mismatches 302; Indels 0; Gaps 0;

Qy   673 GGAAGAGGTGAAGAACAACGGATGAATCGAGTGGACCAAAGAATGCAACAAAGA 732
            || || ||| || ||| ||| || |||||

Db   34738 GAAAAGAAGAAGAAGAAGAAGAAGAAGAAGAGAAGAAGAAAGAGAAGAAGAAGAA 34679

Qy   733 GAACTTCAAGAGGAAGATGAAGAAGATGATGAAGATCTTGAGGACGAGGATGTACCCAGAAGA 792
            || || |||| || || | || | ||| || || || ||||||||

Db   34678 GAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGCATCAGAGAAGAGAGAGAAGAAGA 34619

Qy   793 AGAAGTTCGGATGAGGAACCCAAAGTGAAGCAGAGCATCAGAGAAGAGATTAGCCAGG 852
            |||| || |||||||

Db   34618 AGAAGNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNTCAATCATTG 34559

Qy   853 AGATTGAAAAGTAGTCCTAGATTAAAAGAGAGCTTTTACAGAATGCGGAAGTTCAATCATTG 912
            |||||

Db   34558 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 34499

Qy   913 CTCTCTTACCAACCGAATGAGGGATTCTCCACTGAGCAAGGCGAAGGCCTTTGGCTATGAAC 972

Db   34498 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 34439

Qy   973 GATGAGGATGAAAGTGCATTCCGCGCCAATGGAGGCTCGTGCAAAACTAGATCAAAAATCT 1032
            ||||||||||| || | || || |||||

Db   34438 NNNNNAAGGAGGAGGAGGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGGAA 34379

Qy   1033 CAACTTGTGTCTGCTCGGTCTCCATGGTTTTGGAGAGTCTGATGATGAAGACGAAGAAGAT 1092
                                             || | ||||| |||||

Db   34378 GAGGAGGAGGAGGAGGAAGAAGGAGGAGGAGGAGGAAGAAGGAGGAGGAGGAAGGAA 34319

Qy   1093 GAAAA 1097
            |||||

Db   34318 GAAGA 34314

RESULT 5
AAV21209_01/c
Continuation (2 of 17) of AAV21209 from base 100001 (Methanococcus jannaschii circular c
WP   Sequence split into 17 fragments LOCUS AAV21209 Accession Aav21209
WP   Fragment Name   Begin   End
WP   AAV21209_00   1   110000
WP   AAV21209_01   100001   210000
WP   AAV21209_02   200001   310000
WP   AAV21209_03   300001   410000
WP   AAV21209_04   400001   510000
WP   AAV21209_05   500001   610000
WP   AAV21209_06   600001   710000
WP   AAV21209_07   700001   810000
WP   AAV21209_08   800001   910000
WP   AAV21209_09   900001   1010000
WP   AAV21209_10   1000001   1110000
WP   AAV21209_11   1100001   1210000
WP   AAV21209_12   1200001   1310000
WP   AAV21209_13   1300001   1410000
WP   AAV21209_14   1400001   1510000
WP   AAV21209_15   1500001   1610000
WP   AAV21209_16   1600001   1664976

Query Match   2.6%; Score 49.4; DB 2; Length 110000;
Best Local Similarity 49.2%; Pred. No. 0.25;
Matches 158; Conservative 0; Mismatches 161; Indels 2; Gaps 1;

Qy   1160 TTGAAAAATTAAAGTCAAATGATGAATTGAAAAGTGCATTGGACAGAGAATTAAAATATCGAG 1219
            || ||||| |||| || |||| || ||| ||| ||| |||||||

Db   80139 TAGAAAAATCTGATTTACCTGTTAAATATGTAACTGCCTTTTATACTAAAAAAAACTTGGAG 80080

Qy   1220 TTGATGACGTGGAAAAGTATCTTGCTCCAAAGCCGATGAGGAATTCAATCCAAAACCTCAGC 1279
            |||||| |||||||| || |||| ||||| || ||| ||||

Db   80079 AGTATGATTATGCATTAAAATTATTGATAAAATAATTGAAAAAATATCCAAAATCAGCTA 80020

Qy   1280 CTGGGCTACTTTGCTCCACGTAAAATCCCAACAAGACCACGTAAAATGCTTCCATTATTAA 1339

RESULT 7
AAD58258
ID   AAD58258 standard; DNA; 167739 BP.
XX
AC   AAD58258;
XX
DT   20-NOV-2003 (first entry)
XX
DE   Murine tumour suppressor gene, Lmt reverse complement DNA.
XX
KW   Tumour suppressor gene; Lmt; cancer; therapy; cytostatic; murine; ds.
XX
OS   Mus sp.
XX
PN   WO2003066869-A1.
XX
PD   14-AUG-2003.
XX
PF   07-FEB-2003; 2003WO-AU000126.
XX
PR   07-FEB-2002; 2002AU-00000371.
XX
PA   (HALL-) HALL INST MEDICAL RES WALTER & ELIZA.
XX
PI   Cook WD, Mccaw BJ;
XX
DR   WPI; 2003-646311/61.
XX
PT   New nucleic acid molecule, useful for screening a subject for the
PT   presence of an aberration in a gene encoding an LMT.
XX
PS   Claim 10; Page 127; 373pp; English.
XX
CC   The invention relates to novel tumour suppressor gene, referred to as
CC   Lmt. The invention also relates to the field of cancer therapy and cancer
CC   diagnostics. The nucleic acid molecule is useful for screening a subject
CC   for the presence of an aberration in a gene encoding an LMT. The present
CC   sequence is murine Lmt reverse complement DNA
XX
SQ   Sequence 167739 BP; 44122 A; 33199 C; 34737 G; 45579 T; 0 U; 10102 Other;

Query Match   2.5%; Score 48.2; DB 9; Length 167739;
Best Local Similarity 55.8%; Pred. No. 0.62;
Matches   92; Conservative   0; Mismatches   73; Indels   0; Gaps   0;

Qy   717 AAGAATGCAACAAAGAACTTCAAGAGGAAGATGAAGATGATGAGATCTTGAGGACGA 776
              ||||||||| ||| ||||| |||| ||||||| ||| || |||| |||| ||||
Db   106319 AAGGAAGGAAGGAAGGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGA 106378

Qy   777 GGATGTACCCAGAAGAAGAAGTTCGGATGGAGAACCCACAAAGTGAAGCAGAGCATCAGAG 836
              |||||||| ||||||||||| ||||
Db   106379 AGAAGAAGAAGAAGAAGAAGAGAGAGAAAAGAGAGACAAAGAAAAGGAAGGAAGGAAGAAG 106438

Qy   837 AAGAGATTTAGCCAGGAGATTGAAAAGTAGTCCTAGATTAAAAGA 881
              ||||| ||| ||| || || |||||
Db   106439 GAAGGAAGAAAGAAAGAAGAAGAAGAAAGAAAGAAAGAAAGAAGA 106483

RESULT 8
ABD33029
ID   ABD33029 standard; DNA; 13407 BP.
XX
AC   ABD33029;
XX
DT   18-NOV-2004 (first entry)
XX
DE   Mouse cancer-associated genomic DNA MD22-016.
XX
KW   Mouse; ds; cancer-associated protein; gene; cytostatic; cancer;
KW   leukaemia; lymphoma; CAP.
XX
OS   Mus musculus.

XX
PN   WO2004074320-A2.
XX
PD   02-SEP-2004.
XX
PF   17-FEB-2004; 2004WO-US004730.
XX
PR   14-FEB-2003; 2003US-00367094.
PR   14-MAR-2003; 2003US-00388838.
PR   15-APR-2003; 2003US-00417375.
PR   13-JUN-2003; 2003US-00461862.
PR   15-SEP-2003; 2003US-00663431.
PR   15-DEC-2003; 2003US-00737318.
XX
PA   (SAGR-) SAGRES DISCOVERY INC.
XX
PI   Morris DW, Morris DW, Malandro MS;
XX
DR   WPI; 2004-652914/63.
XX
PT   New isolated cancer-associated polynucleotides and polypeptides useful
PT   for diagnosing, preventing or treating cancers, especially lymphoma and
PT   leukemia, or in screening for agents that modulate cancer.
XX
PS   disclosure; seqid 817; 310pp; English.
XX
CC   The invention relates to an isolated nucleic acid comprising at least 10
CC   contiguous nucleotides of any of the 233 polynucleotide sequences given
CC   in the specification, or its complement. The nucleic acids encode cancer-
CC   associated proteins. Also included are an expression vector comprising
CC   the isolated nucleic acid cited above, a host cell comprising the above
CC   recombinant nucleic acid or expression vector, a microarray for detecting
CC   a cancer-associated (CA) nucleic acid comprising at least one probe
CC   comprising at least 10 contiguous nucleotides of any of the above-
CC   mentioned nucleotide sequences, an isolated polypeptide (encoded within
CC   an open reading frame of a CA sequence selected from any of the 95
CC   polynucleotide sequences as mentioned in the specification, or its
CC   complement), an isolated antibody, (or its antigen binding fragment) that
CC   binds to the above polypeptide, a hybridoma that produces the above
CC   monoclonal antibody, a pharmaceutical composition comprising the above
CC   antibody and a pharmaceutical excipient, a kit for detecting cancer
CC   cells(comprising the antibody cited above, methods for diagnosing cancer
CC   or for detecting the presence or absence of cancer cells in an
CC   individual, a method for inhibiting growth of cancer cells in an
CC   individual, a method for delivering a therapeutic agent to cancer cells
CC   in an individual, an electronic library comprising the above
CC   polynucleotide or polypeptide (or their fragments), methods of screening
CC   for anticancer activity or for a bioactive agent capable of modulating
CC   the activity of a CA protein (CAP), methods for detecting cancer
CC   associated with expression of a polypeptide in a test cell sample, a
CC   method for treating cancers and a method for inhibiting the expression of
CC   CA gene in a cell. The composition and methods are useful for detecting,
CC   diagnosing, preventing and treating cancers, especially lymphoma and
CC   leukaemia. These may also be used in screening for agents that modulate
CC   cancer. The present sequence is a mouse CAP genomic sequence. Note: The
CC   sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 13407 BP; 3129 A; 3381 C; 3740 G; 3157 T; 0 U; 0 Other;

Query Match   2.5%;   Score 47.8; DB 13; Length 13407;
Best Local Similarity 48.4%;   Pred. No. 0.27;
Matches 133; Conservative   0; Mismatches 142; Indels   0; Gaps   0;

Qy   648 AGGTAAAAATGCCCCAGTTGCCGGTGGAAGAGGTGAAGAACAACGGATGATGATGAATCG 707
              ||||||| || ||| || ||||||||||||||||||||| |||||||||||
Db   7975 AGAAGAAGGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGGAAGGAGAA 8034

Qy   708 AGTTGGACCAAAGAATGCAACAACAAAGAACTTCAAGAGAGAAGAGGAAGATGAAGATGATCT 767
              || |||| |||| ||| ||| || ||| |||| || |||||| |||||||||
Db   8035 GAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGA 8094

```
Qy    768 TGAGGACGAGGATGTACCCAGAGAAGAAGTTCGGATGGAGAACCACAAAGTGAAGCAGA 827
          ||||    |||| |   ||  |||| | |  |  |   || |||   ||||||||||||||
Db   8095 AGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAAGAAGAAGAAGAAGAAGAAGAAGAAGAGA 8154

Qy    828 GCATCAGAGAAGAGAGATTTAGCCAGGAGATTGAAAAGTAGTCCTAGATTAAAAGAGCTTTT 887
          || | |||  | ||||  ||| || | ||||         ||||
Db   8155 AGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAAGAAGAAGAAGAAGAAGAAGAAGAAGATGATTT 8214

Qy    888 ACAGAATGCCGAAGTTCAATCATTGCTCTCTTACC 922
          || | ||| | |  ||  ||| | ||| | |||||
Db   8215 CCACAATGAGGGAGCTCAACACATTATCTACCTTCCC 8249
```

RESULT 9
ADL13931/c
ID   ADL13931 standard; DNA; 180385 BP.
XX
AC   ADL13931;
XX
DT   06-MAY-2004 (first entry)
XX
DE   Osteoarthritis-associated polymorphic nucleotide #463.
XX
KW   ds; gene; osteopathic; antiinflammatory; antiarthritic; gene therapy;
KW   joint space narrowing; osteophyte development; joint pain;
KW   osteoarthritis; SNP; single nucleotide polymorphism.
XX
OS   Homo sapiens.
XX
PN   WO2003054166-A2.
XX
PD   03-JUL-2003.
XX
PF   19-DEC-2002; 2002WO-US041225.
XX
PR   20-DEC-2001; 2001US-0342603P.
XX
PA   (INCY-) INCYTE GENOMICS INC.
XX
PI   Jones KA, Schafer A;
XX
DR   WPI; 2003-559141/52.
XX
PT   Determining susceptibility of an individual to joint space narrowing,
PT   osteophyte development and/or joint pain comprises identifying whether
PT   the individual has at least one polymorphism in a polynucleotide encoding
PT   a protein.
XX
PS   Disclosure; SEQ ID NO 463; 297pp; English.
XX
CC   The invention relates to a method of determining susceptibility of an
CC   individual to joint space narrowing and/or osteophyte development and/or
CC   joint pain comprising identifying whether the individual has at least one
CC   polymorphism in a polynucleotide encoding at least one of the protein
CC   listed in the specification. The methods, composition and agent are
CC   useful for modulating the susceptibility of an individual to joint space
CC   narrowing and/or osteophyte development and/or joint pain that is
CC   associated with a disease, preferably osteoarthritis. The cell line and
CC   the non-human animal are useful for screening for an agent for diagnosing
CC   an individual having susceptibility to joint space narrowing and/or
CC   osteophyte development and/or joint pain. This sequence corresponds to
CC   the polynucleotide encoding a protein listed in the specification. (Note:
CC   The sequence data for this patent did not form part of the printed
CC   specification but was obtained in electronic format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences).
XX
SQ   Sequence 180385 BP; 49224 A; 37659 C; 40637 G; 52865 T; 0 U; 0 Other;

Query Match   2.5%; Score 47.8; DB 10; Length 180385;
Best Local Similarity 52.7%; Pred. No. 0.8;
Matches 126; Conservative   0; Mismatches 112; Indels   1; Gaps   1;

```
Qy    635 AAAGAGCAATGTTAGGTAAAAATGCCCAGTTGCCGGTGGAAGAGGTGAAGAACAACGGA 694
```

```
Db  38603 AAAGTAAAAAATAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAG 38544
          ||||     |  || |||| | || ||| |  |||| |||| ||  || |||||||||||
Qy    695 TGATGATGAATCGAGTGGACCAAAGAATGCACAAAGAGAACTTCAAGAGGAAGATGAAG 754
             |||  || || | ||| |  || | || |||| | |  |  ||| | |||||||||||
Db  38543 GAAAAGGAGGAGGAGGAGGAGGAGGATCGTACCCAGAGAAGAAGAAGTTCCGATGGAGAACCAC 38484
          ||||    |||    |  |  |  ||| | ||  |||| |||| ||||||||||||
Qy    755 ATGATGATCTTGAGGACGAGGATGTACCCAGAGAAGAAGAAGTTCCGATGGAGAACCAC 814
              ||| |||  ||    ||  |||| |||| |||| |||||| |||| || |||||||
Db  38483 AAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAACGAG 38424
          |  ||  || |  ||  || | | |||||   |||  |   |  |||   | |||||||
Qy    815 AAGTGAAGCAGAGCATCAGAGAAGAAGAGATTTAGCCAGGAGATTGAAAAGTAGTCCTAGA 873
             |||| |   |  |||| | ||| |||| | |   ||| | || || | ||| || |||
Db  38423 AAGGAGGAGGAGGAAGAAGA-AGAAGAAGAAGAAGAACGAGAAGAAGAAGAAGAAGAAGAGA 38366
```

RESULT 10
ACA28380
ID   ACA28380 standard; DNA; 3534 BP.
XX
AC   ACA28380;
XX
DT   19-JUN-2003 (first entry)
XX
DE   Prokaryotic essential gene #10037.
XX
KW   Antisense; ds; prokaryotic essential gene; cell proliferation;
KW   drug design; gene.
XX
OS   Clostridium botulinum.
XX
PN   WO2020277183-A2.
XX
PD   03-OCT-2002.
XX
PF   21-MAR-2002; 2002WO-US009107.
XX
PR   21-MAR-2001; 2001US-00815242.
PR   06-SEP-2001; 2001US-00948993.
PR   25-OCT-2001; 2001US-0342923P.
PR   08-FEB-2002; 2002US-00072851.
PR   06-MAR-2002; 2002US-0362699P.
XX
PA   (ELIT-) ELITRA PHARM INC.
XX
PI   Wang L,   Zamudio C,   Malone C,   Haselbeck R,   Ohlsen KL,   Zyskind JW;
PI   Wall D,   Trawick JD,   Carr GJ,   Yamamoto R,   Forsyth RA,   Xu HH;
XX
DR   WPI; 2003-029926/02.
DR   P-PSDB; ABU24510.
XX
PT   New antisense nucleic acids, useful for identifying proteins or screening
PT   for homologous nucleic acids required for cellular proliferation to
PT   isolate candidate molecules for rational drug discovery programs.
XX
PS   Claim 14; SEQ ID NO 16250; 1766pp; English.
XX
CC   The invention relates to an isolated nucleic acid comprising any one of
CC   the 6213 antisense sequences given in the specification where expression
CC   of the nucleic acid inhibits proliferation of a cell. Also included are:
CC   (1) a vector comprising a promoter operably linked to the nucleic acid
CC   encoding a polypeptide whose expression is inhibited by the antisense
CC   nucleic acid; (2) a host cell containing the vector; (3) an isolated
CC   polypeptide or its fragment whose expression is inhibited by the
CC   antisense nucleic acid; (4) an antibody capable of specifically binding
CC   the polypeptide;(5) producing the polypeptide; (6) inhibiting cellular
CC   proliferation or the activity of a gene in an operon required for
CC   proliferation; (7) identifying a compound that influences the activity of
CC   the gene product or that has an activity against a biological pathway
CC   required for proliferation, or that inhibits cellular proliferation; (8)
CC   identifying a gene required for cellular proliferation or the biological
CC   pathway in which a proliferation-required gene or its gene product lies
CC   or a gene on which the test compound that inhibits proliferation of an

```
CC    organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC    compound's activity; (11) a culture comprising strains in which the gene
CC    product is overexpressed or underexpressed; (12) determining the extent
CC    to which each of the strains is present in a culture or collection of
CC    strains; or (13) identifying the target of a compound that inhibits the
CC    proliferation of an organism. The antisense nucleic acids are useful for
CC    identifying proteins or screening for homologous nucleic acids required
CC    for cellular proliferation to isolate candidate molecules for rational
CC    drug discovery programs, or for screening homologous nucleic acids
CC    required for proliferation in cells other than S. aureus, S. typhimurium,
CC    K. pneumoniae or P. aeruginosa. The present sequence is one of the target
CC    prokaryotic essential genes. Note: The sequence data for this patent did
CC    not form part of the printed specification, but was obtained in
CC    electronic format directly from WIPO at
CC    ftp.wipo.int/pub/published_pct_sequences
XX
SQ    Sequence 3534 BP; 1742 A; 253 C; 655 G; 884 T; 0 U; 0 Other;

Query Match    2.4%;   Score 47.2;   DB 8;   Length 3534;
Best Local Similarity   46.0%;   Pred. No. 0.22;
Matches  160;  Conservative    0;  Mismatches  188;  Indels    0;  Gaps    0;

Qy    1174  TCAAATGATGAATTGAAAAGTGCATTGGACAGAATTAAATATCGAGTTGATGACGTGGAA 1233
Db    1204  TCTAAGGAAGTAAAATTAAAAGAAGAAGAAATAAATTCATTAAAAATAGATGGAGACAGA 1263

Qy    1234  AAGTATCTTGCTCCAAAGCCGATGGAATTCAATCCAAAACCTCCAGCCTGGCTACTTTGCT 1293
Db    1264  AGAGAAAAGATACAAAACTATATGAAAACTATAAAGAATATAAAAGATTAAATTTAGAA 1323

Qy    1294  CCACGTAAAATCCCAACAAGACCACGTAAAATGCTTCCATTATTAATTGGATCTGATCCA 1353
Db    1324  GTTTGCAATATAACAGAAGAATAAATATAAAATATAAAAGAGCTTAAAAGAGTATGAATTA 1383

Qy    1354  AAAGTTCAAGAGGAAATCGAAGACATCCAAGTACCGAATGCAAAAGAATCA 1413
Db    1384  AAGTATAAAGATACATTAGAAACCACAAAGAGATATAAAAAAATTAGAATATGTTTCA 1443

Qy    1414  AGAGTTTTGACAAATTTGAAGAATAAATCCAAGTCTTGCTGCATTGTTCATGGATGATAAA 1473
Db    1444  GGCAAAAGAAGAAGATTTAATAAAAAAACTCTCCTGGTAACAATGATATTTTGCTTGATAAA 1503

Qy    1474  TTAGAGAATACATTGAAAGGAAGGCAAATGTTAACTGATGAACAGAAA 1521
Db    1504  AAAGATTATATAAATAAATTAATGGAAATTTTTAATGATTTAAATAA 1551

RESULT 11
AAQ87587
ID    AAQ87587 standard; DNA; 1686 BP.
XX
AC    AAQ87587;
XX
DT    27-AUG-2003  (revised)
DT    19-DEC-1995  (first entry)
XX
DE    DNA encoding Leucocytozoan protozoa structural protein epitope.
XX
KW    leucocytozoan protozoa; structural protein; epitope; vaccine; fowl;
KW    leucocytozoanosis; treatment; ss.
XX
OS    Leucocytozoon.
XX
PN    JP07089995-A.
XX
PD    04-APR-1995.
XX
PF    10-SEP-1993;   93JP-00226078.
XX
PR    10-SEP-1993;   93JP-00226078.
XX
PA    (DOBU-) DOBUTSUYO SEIBUTSUGAKUTEKI SEIZAI KYOKAI.
PA    (NISS-) NISSEIKEN KK.
```

```
XX
DR    WPI; 1995-167252/22.
DR    P-PSDB; AAR70491.
XX
PT    Immune inducing polypeptide against Leucocytozoan protozoa - useful in
PT    production of vaccines for treatment of leucocytozoanosis in fowl.
XX
PS    Claim 1; Page 12-14; 20pp; Japanese.
XX
CC    AAQ87587-89 encode polypeptides having a whole or partial epitope of a
CC    structural protein of Leucocytozoan protozoa (see AAR70491-93). The
CC    polypeptides and DNA encoding them are useful in the production of
CC    vaccines for the treatment of leucocytozoanosis of fowl. (Updated on 27-
CC    AUG-2003 to correct OS field.)
XX
SQ    Sequence 1686 BP; 915 A; 129 C; 382 G; 260 T; 0 U; 0 Other;

Query Match    2.4%;   Score 47;   DB 2;   Length 1686;
Best Local Similarity   44.2%;   Pred. No. 0.18;
Matches  194;  Conservative    0;  Mismatches  245;  Indels    0;  Gaps    0;

Qy    652   AAAAATGCCCAGTTGCCGTGGAAGAGGTGAAGAACAACGATGATGAATCGAGTG 711
Db    124   AAAGAAGAACAAGAAGAAGAAGACAAGAACAAGAACAAGAAAATCGTAGAA 183

Qy    712   GACCAAAGAATGCAACAAAGAGAACTTCAAGAGAGAAGATGAACGATGATCTTGAG 771
Db    184   GAACAAGAACAAGATGAACGAAGAACAAGAACAAGAAGATGAAGGAGAAGAAGAAAAA 243

Qy    772   GACGAGGAGGATGACCCAGGAAGAAGAAGTTCGGATGGAGAAGAAGAAGAAGAAGAAGAGCAT 831
Db    244   GAAGAAGAAGAAGAAGAAGAAGAAGAAGACCACAAAGTGAAGGAGAGAGCA 303

Qy    832   CAGAGAAGAAGAGATTTAGCCAGGAGGATTGAAAAGTAGTCCTAGATTAAAAGAGCTTTTACAG 891
Db    304   GATGAAGAAGAAGAAGAAGAAGAACAAGATGAAGAAGAAGAAGATGAA 363

Qy    892   AATGCGGAAGTTCAATCATTGCTCTCTTACCACGATGAGGGATTCTCCACTGAGCAAG 951
Db    364   GAAGAAGAGAAGAAGAAGAACAAAATGAAGATGAAGATGAACAAAATGAA 423

Qy    952   CGAAGGCCTTTGCCTATGAACGACCGATGAGGATGAAGAAGAAGAAAGTGCATTCCGCGCAATGGCGAGGCTCGT 1011
Db    424   GATGAACAAAATGAAGAAGAAGAAGAACAAGAACAAGATGAA 483

Qy    1012  GCAAAACTAGATCAAAAATCTCAACTTGTCTCCGGTCTCCGGGAAGAAGAAGAAGAAGAAGAACAA 543
Db    484   GAAGAACAAGATGAAGAAGAACAAGATGAAGAAGATGAAGAAGAAGAACAA

Qy    1072  GATGATGAAGACGAAGAAG 1090
Db    544   GATGAAGAACAAGAAGAAG 562
```

```
RESULT 12
AAS88862/c
ID    AAS88862 standard; cDNA; 2427 BP.
XX
AC    AAS88862;
XX
DT    13-FEB-2002  (first entry)
XX
DE    DNA encoding novel human diagnostic protein #24666.
XX
KW    Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW    food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS    Homo sapiens.
XX
PN    WO20017S067-A2.
XX
PD    11-OCT-2001.
XX
```

PF   30-MAR-2001; 2001WO-US008631.
XX
PR   31-MAR-2000; 2000US-00540217.
PR   23-AUG-2000; 2000US-00649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT, Liu C, Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   P-PSDB; ABG24675.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity.
XX
PS   Claim 1; SEQ ID NO 24666; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC   sequences. (I) is useful as hybridisation probes, polymerase chain
CC   reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC   and in recombinant production of (II). The polynucleotides are also used
CC   in diagnostics as expressed sequence tags for identifying expressed
CC   genes. (I) is useful in gene therapy techniques to restore normal
CC   activity of (II) or to treat disease states involving (II). (II) is
CC   useful for generating antibodies against it, detecting or quantitating a
CC   polypeptide in tissue, as molecular weight markers and as a food
CC   supplement. (II) and its binding partners are useful in medical imaging
CC   of sites expressing (II). (I) and (II) are useful for treating disorders
CC   involving aberrant protein expression or biological activity. The
CC   polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic
CC   coding sequences of the invention. Note: The sequence data for this
CC   patent did not appear in the printed specification, but was obtained in
CC   electronic format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 2427 BP; 467 A; 680 C; 443 G; 837 T; 0 U; 0 Other;

Query Match   2.4%;   Score 46.2; DB 5; Length 2427;
Best Local Similarity 52.9%;   Pred. No. 0.35;
Matches   99;   Conservative   0; Mismatches   88; Indels   0; Gaps   0;

Qy   693 GATGATGATGATCGAGTGGACCAAAGAATGCAACAAAGAGAACTTCAAGAGGAAGATGA 752
Db   1788 GAGGAAGAGGAAGAGGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAACTTCAAGAGGAAGATGA 1729

Qy   753 AGATGATGATGATCTTGAGGACGACGAGGATGTACCCAGAAGAGGAAGAAGAAGAAGAAGA 812
Db   1728 AGAAGAAGAACAAGAACAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAG 1669

Qy   813 ACAAAGTGAAGCAGAGCCATCAGAAGAGAGATTTAGCCAGGAGATTGAAAAGTAGTCCTAG 872
Db   1668 AAGAAGAACAAGAACAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAG 1609

Qy   873 ATTAAAA 879
Db   1608 AAGAAGA 1602

RESULT 13
AAS69554
ID   AAS69554 standard; cDNA; 2427 BP.
XX
AC   AAS69554;
XX
DT   13-FEB-2002 (first entry)
XX
DE   DNA encoding novel human diagnostic protein #5358.

XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US008631.
XX
PR   31-MAR-2000; 2000US-00540217.
PR   23-AUG-2000; 2000US-00649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT, Liu C, Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   P-PSDB; ABG05367.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity.
XX
PS   Claim 1; SEQ ID NO 5358; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC   sequences. (I) is useful as hybridisation probes, polymerase chain
CC   reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC   and in recombinant production of (II). The polynucleotides are also used
CC   in diagnostics as expressed sequence tags for identifying expressed
CC   genes. (I) is useful in gene therapy techniques to restore normal
CC   activity of (II) or to treat disease states involving (II). (II) is
CC   useful for generating antibodies against it, detecting or quantitating a
CC   polypeptide in tissue, as molecular weight markers and as a food
CC   supplement. (II) and its binding partners are useful in medical imaging
CC   of sites expressing (II). (I) and (II) are useful for treating disorders
CC   involving aberrant protein expression or biological activity. The
CC   polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic
CC   coding sequences of the invention. Note: The sequence data for this
CC   patent did not appear in the printed specification, but was obtained in
CC   electronic format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 2427 BP; 837 A; 443 C; 680 G; 467 T; 0 U; 0 Other;

Query Match   2.4%;   Score 46.2; DB 5; Length 2427;
Best Local Similarity 52.9%;   Pred. No. 0.35;
Matches   99;   Conservative   0; Mismatches   88; Indels   0; Gaps   0;

Qy   693 GATGATGATGATCGAGTGGACCAAAGAATGCAACAAAGAGAACTTCAAGAGGAAGATGA 752
Db   640 GAGGAAGAGGAAGAGGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGTTCGGATGGAGAACC 699

Qy   753 AGATGATGATGATCTTGAGGACGACGAGGATGTACCCAGAAGAGGAAGAAGAAGAAGAAGA 812
Db   700 AGAAGAAGAACAAGAACAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAG 759

Qy   813 ACAAAGTGAAGCAGAGCCATCAGAAGAGAGATTTAGCCAGGAGATTGAAAAGTAGTCCTAG 872
Db   760 AAGAAGAACAAGAACAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAG 819

Qy   873 ATTAAAA 879
Db   820 AAGAAGA 826

RESULT 14
AAS90685
ID   AAS90685 standard; cDNA; 423 BP.
XX
AC   AAS90685;
XX
DT   13-FEB-2002 (first entry)
XX
DE   DNA encoding novel human diagnostic protein #26489.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US008631.
XX
PR   31-MAR-2000; 2000US-00540217.
PR   23-AUG-2000; 2000US-00649167.
XX
PA   (HYSB-) HYSEQ INC.
XX
PI   Drmanac RT, Liu C,   Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   P-PSDB; ABG26498.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity.
XX
PS   Claim 1; SEQ ID NO 26489; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC   sequences. (I) is useful as hybridisation probes, polymerase chain
CC   reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC   and in recombinant production of (II). The polynucleotides are also used
CC   in diagnostics as expressed sequence tags for identifying expressed
CC   genes. (I) is useful in gene therapy techniques to restore normal
CC   activity of (II) or to treat disease states involving (II). (II) is
CC   useful for generating antibodies against it, detecting or quantitating a
CC   polypeptide in tissue, as molecular weight markers and as a food
CC   supplement. (II) and its binding partners are useful in medical imaging
CC   of sites expressing (II). (I) and (II) are useful for treating disorders
CC   involving aberrant protein expression or biological activity. The
CC   polypeptide and polynucleotide sequences have application in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic
CC   coding sequences of the invention. Note: The sequence data for this
CC   patent did not appear in the printed specification, but was obtained in
CC   electronic format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 423 BP; 235 A; 11 C; 159 G; 18 T; 0 U; 0 Other;

Query Match   2.4%;   Score 46; DB 5; Length 423;
Best Local Similarity 51.2%;   Pred. No. 0.19;
Matches 132;   Conservative   0;   Mismatches 125;   Indels   1;   Gaps   1;

Qy   648 AGGTAAAAATGCCCCAGTTGCCCGTGGAAGAGGTGAAGAACAACGATGATGAATCG 707
             |||| | || | | ||||||||||||||||||||||||||||||||||
Db   167 AGAAGAAAAGGAAGAAGAAGAAGAAGAAGAAGGAAGAAGAAGAACAAGAGAAGAAG 226

Qy   708 AGTGGACCAAAGATGCAACAAAGAGAACTTCAAGAGAGAGAAGAGAAGATGATGATCT 767
             | |||| ||| ||| ||| |||||||||||| ||||||| || || ||

Db   227 AAGAAGTCAAATATCTGGAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAGGAGGA 286
             |||| | ||||||||| || || || || || ||||||| |||| |
Qy   768 TGACGACGAGGCGGTACCCAGAAGAAGAAGAGTTCGGATGCAGGAAGTGAAGCAGA 827
              |||| || ||| ||||||||| ||| || ||||||| ||||| ||||| |
Db   287 GGAAGAAGGAGAAGGAGAAGGAGAAGGAGAAGGAGAAGGAGAAGGAGAAGAAGAGA 346
              ||| || |||| || || || |||||| || ||| || || ||| ||||| |
Qy   828 GCATCAGAGAAGAGATTTAGCCAGGAGATTGAAAAGTAGTCCTAGATTAAAAGAGCTTTT 887
             | ||| |||||| |||||||| ||| | || || | |||| ||| |||| |
Db   347 AGAGA-AGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAAATCAAATA 405
             ||||| |||||||| | |||| ||||

Qy   888 ACAGAATGCGGAAGTTCA 905
             ||||||||| ||| | |
Db   406 TCTGGAGCCCGGATTTGA 423

RESULT 15
AAS45323/c
ID   AAS45323 standard; DNA; 17848 BP.
XX
AC   AAS45323;
XX
DT   18-DEC-2001 (first entry)
XX
DE   Chemically pretreated complementary DNA associated with cell cycle #14.
XX
KW   Cell cycle; human; CpG dinucleotide; cytosine methylation; HIV; aging;
KW   human immunodeficiency virus; neurodegenerative disorder; solid tumour;
KW   graft-versus-host disease; glomerular disease; Lewy body disease; cancer;
KW   arthritis; arteriosclerosis; anti-HIV; neuroprotective; antiarthritic;
KW   immunosuppressive; antitumour; cytostatic; antiarteriosclerotic; ds;
KW   PCR primer.
XX
OS   Homo sapiens.
XX
PN   WO200168911-A2.
XX
PD   20-SEP-2001.
XX
PF   15-MAR-2001; 2001WO-EP002945.
XX
PR   15-MAR-2000; 2000DE-01013847.
PR   06-APR-2000; 2000DE-01019058.
PR   07-APR-2000; 2000DE-01019173.
PR   30-JUN-2000; 2000DE-01032529.
PR   01-SEP-2000; 2000DE-01043826.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Olek A,   Piepenbrock C, Berlin K;
XX
DR   WPI; 2001-602751/68.
XX
PT   Designing primers and probes for analyzing diseases associated with
PT   cytosine methylation state e.g. arthritis, cancer, aging,
PT   arteriosclerosis comprising fragments of chemically modified genes
PT   associated with cell cycle.
XX
PS   Claim 1; SEQ ID NO 28; 28pp; English.
XX
CC   Sequences AAS45296-AAS45520 represent chemically pretreated genomic DNA
CC   molecules associated with the cell cycle and specific PCR primers of the
CC   invention. The sequences are useful for detecting the methylation state
CC   of all CpG dinucleotides in a sequence and therefore for analysing
CC   associated diseases. By analysing cytosine methylations in the pretreated
CC   DNA, genetic and/or epigenetic parameters for the diagnosis and therapy
CC   of existing diseases or the predisposition to specific diseases can be
CC   ascertained. The parameters may be compared to another set of genetic
CC   and/or epigenetic parameters, the differences serving as basis for
CC   diagnosis and/or prognosis events which are disadvantageous to patients.
CC   The sequences of the invention are useful for the diagnosis and therapy
CC   of HIV infection, neurodegenerative disorders, graft-versus-host disease,
CC   aging, glomerular disease, Lewy body disease, arthritis,
CC   arteriosclerosis, solid tumours and cancers

XX
SQ   Sequence 17848 BP; 5055 A; 211 C; 3533 G; 9043 T; 0 U; 6 Other;

Query Match   2.4%;   Score 45.8;   DB 4;   Length 17848;
Best Local Similarity   48.6%;   Pred. No. 1;
Matches 125;   Conservative   0;   Mismatches 132;   Indels   0;   Gaps   0;

Qy   1340 TTGGATCTGATCCAAAAGTTCAAGAGGAAATACGAAGCATCCAAGTACCGAATGGAAAA 1399
           ||||||||| |||| || |||| ||||||| ||||||||| || ||| |||| |||
Db   7119 TTATAACTATTAAAATAATTAAAAAAAAAAAATTACAAACCTAAATAAAAAACTAAAA 7060

Qy   1400 TTGCAAAAGAATCAAGAGTTTTGACAAATTTGAAGAATAATCCAAGTCTTGCTGCATTGT 1459
           | |||||| |||| ||| || |||||| | ||||||| || || ||
Db   7059 TAACAAAAAATAACGATTTCAAAAAAATTTTAAAAAACGAAATTAACAAAATATAATAAC 7000

Qy   1460 TCATGGATGATAAATTAGAGAATACATTGAAAGGAGGCAAATGTTAACTGATGAACAGA 1519
           | |||||| |||| ||| |||| || |||| ||||| || |||||||
Db   6999 CACTTAAAATAAAATAAAAAAAAAAAAATAAAAAATAACACTAAAATTCATAAAAATA 6940

Qy   1520 AAGGTAGAACCACGTGTCAAAAACAATTCGTGCATTACCAAGACTGTTCGGTGCACCAACTG 1579
           |||| || ||| || | | | ||| || ||| |
Db   6939 AAAATAAAAAAAAACGATCTAAAAAAATTAAAATAAATAATAATTTATATAAAATACCTA 6880

Qy   1580 CAAAAGCTGAAATGATT 1596
           ||||| ||| ||| |||
Db   6879 TAAAACATAAAAATATT 6863

Search completed: July 15, 2005, 15:15:32
Job time : 1078 secs

```
                    GenCore version 5.1.6
         Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on:    July 15, 2005, 10:39:50 ; Search time 340 Seconds
                                     (without alignments)
                          9297.902 Million cell updates/sec

Title:          US-10-736-868-1
Perfect score:  1932
Sequence:       1 atgatttattttatttt.........tgcctgcactgggattgtag 1932

Scoring table:  IDENTITY NUC
                Gapop 10.0 , Gapext 1.0

Searched:       1202784 seqs, 81813359 residues

Total number of hits satisfying chosen parameters:    2405568

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :      Issued Patents NA:*
                1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
                2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
                3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
                4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
                5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
                6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.
```

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
c 1	75.4	3.9	7218	1	US-08-232-463-14	Sequence 14, Appl
c 2	49.4	2.6	1664976	1	US-08-916-421B-1	Sequence 1, Appli
c 3	49.4	2.6	1664976	4	US-09-692-570-1	Sequence 1, Appli
4	47.6	2.5	1141	4	US-09-806-708B-22	Sequence 22, Appl
5	46.8	2.4	1086	4	US-09-248-796A-6002	Sequence 6002, Ap
6	46	2.4	289	3	US-09-007-005-17	Sequence 17, Appl
7	46	2.4	289	3	US-09-244-796-17	Sequence 17, Appl
c 8	45.8	2.4	247781	4	US-09-949-016-14193	Sequence 14193, A
c 9	44.2	2.3	601	4	US-09-949-016-86857	Sequence 86857, A
c 10	43.8	2.3	64309	4	US-09-949-016-14581	Sequence 14581, A
c 11	43.6	2.3	601	4	US-09-949-016-91474	Sequence 91474, A
c 12	43.6	2.3	601	4	US-09-949-016-91475	Sequence 91475, A
13	43.6	2.3	107980	4	US-09-949-016-14370	Sequence 14370, A
14	43.2	2.2	1755	4	US-09-248-796A-3448	Sequence 3448, Ap
15	42.2	2.2	213456	4	US-09-820-007-3	Sequence 3, Appli
16	42.2	2.2	323820	4	US-09-949-016-14139	Sequence 14139, A
c 17	41.6	2.2	1671	4	US-09-248-796A-4970	Sequence 4970, Ap
18	41.2	2.1	696	3	US-09-461-697-193	Sequence 193, App
19	41.2	2.1	699	3	US-09-461-697-191	Sequence 191, App
20	41.2	2.1	717	3	US-09-461-697-189	Sequence 189, App
21	41.2	2.1	774	3	US-09-461-697-187	Sequence 187, App
22	41.2	2.1	819	3	US-09-461-697-185	Sequence 185, App
23	41.2	2.1	1669	3	US-09-461-697-184	Sequence 184, App
24	41.2	2.1	3095	6	5231168-1	Patent No. 5231168
25	41	2.1	3095	6	5231168-1	Patent No. 5231168
c 26	41	2.1	150394	4	US-09-949-016-13042	Sequence 13042, A
c 27	40.8	2.1	832	4	US-09-621-976-2813	Sequence 2813, Ap
c 28	40.8	2.1	929	4	US-09-671-317-14	Sequence 14, Appl
c 29	40.8	2.1	1001	4	US-09-671-317-439	Sequence 439, App
c 30	40.8	2.1	2119	4	US-09-399-588-1	Sequence 1, Appli
c 31	40.6	2.1	495	4	US-09-248-796A-759	Sequence 759, App
c 32	40.6	2.1	1141	4	US-09-806-708B-22	Sequence 22, Appl
33	40.6	2.1	43657	4	US-09-949-016-13777	Sequence 13777, A
34	39.8	2.1	1298	3	US-08-948-705-3	Sequence 3, Appli
35	39.8	2.1	1298	4	US-09-510-543-3	Sequence 3, Appli
36	39.8	2.1	1938	4	US-09-799-451-543	Sequence 543, App
37	39.8	1	4599	1	US-08-431-080-27	Sequence 27, Appl
38	39.8	2.1	4599	3	US-08-938-534-27	Sequence 27, Appl
39	39.8	2.1	4599	3	US-09-345-294-27	Sequence 27, Appl
c 40	39.6	2.0	666	4	US-09-545-216A-3	Sequence 3, Appli
c 41	39.6	2.0	981	4	US-09-545-216A-1	Sequence 1, Appli
42	39.6	2.0	1353	4	US-08-956-171E-521	Sequence 521, App
43	39.6	2.0	1353	4	US-08-781-986A-521	Sequence 521, App
44	39.6	2.0	1893	4	US-09-248-796A-6172	Sequence 6172, Ap
45	39.6	2.0	2822	4	US-09-907-794A-48	Sequence 48, Appl

ALIGNMENTS

```
RESULT 1
US-08-232-463-14/c
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
;  APPLICANT: DORNER, F.
;  APPLICANT: SCHEIFLINGER, F.
;  APPLICANT: FALKNER, F.G.
;  TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
;  NUMBER OF SEQUENCES: 52
;  CORRESPONDENCE ADDRESS:
;  ADDRESSEE: Foley & Lardner
;  STREET: 1800 Diagonal Road, Suite 500
;  CITY: Alexandria
;  STATE: VA
;  COUNTRY: USA
;  ZIP: 22313-0299
;  COMPUTER READABLE FORM:
;  MEDIUM TYPE: Floppy disk
;  COMPUTER: IBM PC compatible
;  OPERATING SYSTEM: PC-DOS/MS-DOS
;  SOFTWARE: PatentIn Release #1.0, Version #1.25
;  CURRENT APPLICATION DATA:
;  APPLICATION NUMBER: US/08/232,463
;  FILING DATE:
;  CLASSIFICATION: 435
;  PRIOR APPLICATION DATA:
;  APPLICATION NUMBER: US/07/935,313
;  FILING DATE:
;  APPLICATION NUMBER: EP 91 114 300.6
;  FILING DATE: 26-AUG-1991
;  ATTORNEY/AGENT INFORMATION:
;  NAME: BENT, Stephen A.
;  REGISTRATION NUMBER: 29,768
;  REFERENCE/DOCKET NUMBER: 30472/114 IMMU
;  TELECOMMUNICATION INFORMATION:
;  TELEPHONE: (703)836-9300
;  TELEFAX: (703)683-4109
;  TELEX: 899149
;  INFORMATION FOR SEQ ID NO: 14:
;  SEQUENCE CHARACTERISTICS:
;  LENGTH: 7218 base pairs
;  TYPE: nucleic acid
;  STRANDEDNESS: single
;  TOPOLOGY: linear
;  IMMEDIATE SOURCE:
;  CLONE: pTZgpt-F1s
US-08-232-463-14

Query Match    3.9%;    Score 75.4;    DB 1;    Length 7218;
```

Best Local Similarity 5.4%; Pred. No. 2.4e-11;
Matches   22; Conservative 237; Mismatches 148; Indels   0; Gaps   0;

```
Qy    622 CAAGCTCTTGCTCAAAGAGCAATGTTAGGTAAAAATGCCCCAGTTGCCGGTGGAAGAGGT 681
              ||||  |||   | ||  :: :::::: :::::::::  : :: ::::::::: :
Db   1466 CAAGTAGTTAAAGAGATAGAGAATTTGGTACRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1407

Qy    682 GAAGAACAACGATGATGATGAATCGAGTCGAGCTGGACCAAAGAATGCAACAAAGAACTTCAA 741
              :::::: :: ::: ::: :::: :::::::: ::  :: :::::::::::
Db   1406 RRRRRRRRRRRRRRRRR RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1347

Qy    742 GAGGAAGATGAAGATGATGATCTTGAGGACGAGGATGTACCCAGAAGAAGAAGTTCG 801
              ::::::::: :::: : : :::::::::::::::: ::
Db   1346 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1287

Qy    802 GATGGAGAACCACAAAGTGAAGCAGAGCATCAGAGAGAGATTTAGCCAGGAGATTGAAA 861
              ::::::::::   ::::::::::: :::::::: ::::  :::::
Db   1286 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1227

Qy    862 AGTAGTCCTAGATTAAAAGAGCTTTTACAGAATGCGGAAGTTCAATCATTGCTCTCTTAC 921
              :::: ::::::::: : ::::::  :: : ::::::::
Db   1226 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1167

Qy    922 CAACGAATGAGGGATTCTCCACTGACCAAGGCGAAGGCCTTTGGCTATGAACGATGAGGAT 981
              :: :::: :: :::::  :::: :::::::::
Db   1166 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1107

Qy    982 GAAAGTGCATTCCGCGCAATGAGAGGCTCGTGCAAAACTAGATCAAAA 1028
              : : :::: :::::     : ::::: :::  :: |||  ||
Db   1106 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRATCGCAA 1060
```

;   NAME/KEY: misc_feature
;   LOCATION: (28222)..(28222)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (28257)..(28258)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (84773)..(84773)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (84808)..(84808)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (84812)..(84812)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (98120)..(98120)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (98159)..(98159)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (98239)..(98239)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (98266)..(98266)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (98343)..(98343)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (103998)..(103998)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (148948)..(148948)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (163385)..(163385)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (191989)..(191989)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (191995)..(191995)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (231980)..(231980)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (234187)..(234187)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (234220)..(234220)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (234814)..(234814)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (309398)..(309398)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (309418)..(309418)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (312837)..(312837)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (312993)..(312993)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (319226)..(319226)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (559167)..(559167)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (559241)..(559241)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (600992)..(600992)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (622708)..(622708)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (657081)..(657081)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (657203)..(657203)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (674435)..(674435)
;   OTHER INFORMATION: n equals a, t, c, or g

```
Qy   1220 TTGATGACGTGAAAAGTATCTTGCTCCAAAGCCGATGGAATTCAATCCAAAACCTCAGC 1279
              ||||    ||      ||     ||||||
Db 180079 AGTATGATTATGCATTAAAAATTATTGATAAAAATATCCAAAATCAGCTA 180020

Qy   1280 CTGGCTACTTTGCTCCACGTAAAATCCCAACAAGACCCACGTAAAATGCTTCCATTATTAA 1339
              |||      |||||      ||    ||    |  ||||
Db 180019 TTGCCTGGGCTGAAAAAGGGGAAATATTATATAGAGAAGCAAACTTAAAAAATCATTAG 179960

Qy   1340 TTGGATCTGATCCAAAGTCAAGAGAGAAATACGAAGACATCCAAGTACCGAATGGAAAA 1399
              ||||||     |||     ||||||     |||
Db 179959 AATGTTTGAT--AATGCTTAAAAATAAATCCCAAAGATTGCCAGTCCTTATTGTACAA 179902

Qy   1400 TTGCAAAGAATCAAGAGTTTGACAAATTTGAAGAATAATCCAAGTCTTGCTGCATTGT 1459
              ||    ||||    ||||||     ||||    |||||
Db 179901 AGGAGAAATATATTATTTAAACTTGGAAGATATGGAAGCATTAAAATGTCTAAAAAAAGT 179842

Qy   1460 TCATGGATGATAAATTAGAGA 1480
              |||    |||   ||||||
Db 179841 TTTTGAAAGAAATAATAAAGA 179821
```

RESULT 3
US-09-692-570-1/c
;   Sequence 1, Application US/09692570
;   Patent No. 6797466
;   GENERAL INFORMATION:
;   APPLICANT: Bult et al.
;   TITLE OF INVENTION: Complete Genome Sequence of the Methanogenic Archaeon, Methanococc
;   Patent No. 6797466
;   TITLE OF INVENTION: jannaschii
;   FILE REFERENCE: PB275C1
;   CURRENT APPLICATION NUMBER: US/09/692,570
;   CURRENT FILING DATE: 2003-01-14
;   PRIOR APPLICATION NUMBER: US 60/024,428
;   PRIOR FILING DATE: 1996-08-22
;   PRIOR APPLICATION NUMBER: US 08/916,421
;   PRIOR FILING DATE: 1997-08-22
;   NUMBER OF SEQ ID NOS: 20
;   SOFTWARE: PatentIn version 3.1
;
;   SEQ ID NO 1
;   LENGTH: 1664976
;   TYPE: DNA
;   ORGANISM: Methanococcus jannaschii
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (28222)..(28222)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (28257)..(28258)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (84773)..(84773)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (84808)..(84808)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (84812)..(84812)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (98120)..(98120)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (98159)..(98159)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
```

```
;   NAME/KEY: misc_feature
;   LOCATION: (682442)..(682442)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (713652)..(713652)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (741684)..(741684)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (779455)..(779455)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (779676)..(779676)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (855539)..(855539)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (871619)..(871619)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (1084830)..(1084830)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (1096846)..(1096846)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (1119881)..(1119881)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (1130881)..(1130881)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (1310988)..(1310988)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (1313224)..(1313224)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (1349473)..(1349473)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (1349491)..(1349491)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (1470091)..(1470091)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (1569020)..(1569020)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (1602912)..(1602912)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (1603734)..(1603734)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (1637998)..(1637998)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (1664854)..(1664855)
;   OTHER INFORMATION: n equals a, t, c, or g
US-08-916-421B-1
```

```
Query Match        2.6%;   Score 49.4;  DB 4;  Length 1664976;
Best Local Similarity  49.2%;  Pred. No. 0.064;
Matches  158;  Conservative  0;  Mismatches 161;  Indels  2;  Gaps  1;

Qy   1160 TTGAGAAATTAAAGTCAAATGATGAATTGAAAAGTGCATTGGACACAGAATTAAATATCGAG 1219
              |  |||||  |||       |||||        ||  || ||     |||||||
Db 180139 TAGAAAAATCTGATTTACCTGTTAAATATGTAACTGCCTTTATACTAAAAAAACTTGGAG 180080
```

LOCATION: (98239)..(98239)
OTHER INFORMATION: n equals a, t, c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (98266)..(98266)
OTHER INFORMATION: n equals a, t, c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (98343)..(98343)
OTHER INFORMATION: n equals a, t, c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (103398)..(103398)
OTHER INFORMATION: n equals a, t, c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (148948)..(148948)
OTHER INFORMATION: n equals a, t, c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (163385)..(163385)
OTHER INFORMATION: n equals a, t, c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (191989)..(191989)
OTHER INFORMATION: n equals a, t, c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (191995)..(191995)
OTHER INFORMATION: n equals a, t, c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (231980)..(231980)
OTHER INFORMATION: n equals a, t, c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (234187)..(234187)
OTHER INFORMATION: n equals a, t, c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (234220)..(234220)
OTHER INFORMATION: n equals a, t, c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (234814)..(234814)
OTHER INFORMATION: n equals a, t, c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (309398)..(309398)
OTHER INFORMATION: n equals a, t, c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (309418)..(309418)
OTHER INFORMATION: n equals a, t, c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (312837)..(312837)
OTHER INFORMATION: n equals a, t, c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (312993)..(312993)
OTHER INFORMATION: n equals a, t, c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (319226)..(319226)
OTHER INFORMATION: n equals a, t, c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (559167)..(559167)
OTHER INFORMATION: n equals a, t, c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (559241)..(559241)

OTHER INFORMATION: n equals a, t, c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (600992)..(600992)
OTHER INFORMATION: n equals a, t, c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (622708)..(622708)
OTHER INFORMATION: n equals a, t, c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (657081)..(657081)
OTHER INFORMATION: n equals a, t, c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (657203)..(657203)
OTHER INFORMATION: n equals a, t, c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (674435)..(674435)
OTHER INFORMATION: n equals a, t, c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (682442)..(682442)
OTHER INFORMATION: n equals a, t, c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (713652)..(713652)
OTHER INFORMATION: n equals a, t, c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (741684)..(741684)
OTHER INFORMATION: n equals a, t, c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (779455)..(779455)
OTHER INFORMATION: n equals a, t, c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (779676)..(779676)
OTHER INFORMATION: n equals a, t, c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (855539)..(855539)
OTHER INFORMATION: n equals a, t, c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (871619)..(871619)
OTHER INFORMATION: n equals a, t, c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1084830)..(1084830)
OTHER INFORMATION: n equals a, t, c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1096846)..(1096846)
OTHER INFORMATION: n equals a, t, c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1119881)..(1119881)
OTHER INFORMATION: n equals a, t, c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1130881)..(1130881)
OTHER INFORMATION: n equals a, t, c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1310988)..(1310988)
OTHER INFORMATION: n equals a, t, c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1313224)..(1313224)
OTHER INFORMATION: n equals a, t, c, or g

```
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1349473)..(1349473)
;   OTHER INFORMATION: n equals a, t, c, or g

Query Match         2.6%;  Score 49.4; DB 4; Length 1664976;
Best Local Similarity  49.2%;  Pred. No. 0.064;
Matches  158; Conservative   0; Mismatches 161; Indels   2; Gaps   1;

Qy    1160 TTGAGAAATAAGTCAAATGATGAATTGAAAAGTGCATTGGACAGAATTAAATATCGAG 1219
                |||||||||||||||||||| ||||||||||    || |||  ||  ||| ||||
Db  180139 TAGAAAAATCTGATTTACCTGTTAAATATGTAACTGCCTTTATACTAAAAAAACTTGGAG 180080

Qy    1220 TTGATGACGTGGAAAAGTATCTTGCTCCAAAGCCGATGGAATTCAATCAAAACCTCAGC 1279
                ||||||||||||||  || |||||||||||||||||||||||||   ||||||||||
Db  180079 AGTATGATTATGCATTAAAAATTATTGATAAAATATCCAAAATATCCAAAATCAGCTA 180020

Qy    1280 CTGCTACTTTGCTCCACGTAAAATCCCAACAGACCACGTAAAATGCTTCCATTATTAA 1339
                || |||||| ||||||||  ||||  ||||||||||  ||||||  ||||| ||||
Db  180019 TTGCCTGGGCTGGAAAAAGGGGAAAATATTTATATAGAGAAGGCAAACTTAAAAAATCATTAG 179960

Qy    1340 TTGGATCTGATCCAAAAGTTCAAGAGAAATACGAAGACATCCAAGTACCGAATGGAAAA 1399
                ||||| ||     ||||||| || ||||||  |||||||||||| || ||||||||
Db  179959 AATGTTTTGAT--AATGCTTAAAAATAAATCCCAAAGATTGCCAGTCCTTATTGTACAA 179902

Qy    1400 TTGCAAAAGAATCAAGAGTTTTGACAAATTTGAAGAATAATCCAAGTCTTGCTGCATTGT 1459
                |||||||||||||||||||||    ||||||||||||||||||||||||||||||||
Db  179901 AGGAGAAATATTATTTAAACTTGGAAGATATGGAGAAGCATTAAAATGTCTAAAAAAAAGT 179842

Qy    1460 TCATGGATGATAAATTAGAGA 1480
                ||||||| |||| ||||||
Db  179841 TTTTGAAAGAAATAATAAAGA 179821
```

RESULT 4
US-09-806-708B-22
; Sequence 22, Application US/09806708B
; Patent No. 6784342
; GENERAL INFORMATION:
; APPLICANT: The University of British Columbia
; TITLE OF INVENTION: Regulation of Embryonic Transcription in Plants
; FILE REFERENCE: 4810-58741
; CURRENT APPLICATION NUMBER: US/09/806,708B
; CURRENT FILING DATE: 2001-04-03
; PRIOR APPLICATION NUMBER: US 60/147,133
; PRIOR FILING DATE: 1999-08-04
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: PatentIn version 3.0
;
; SEQ ID NO 22
;   LENGTH: 1141
;   TYPE: DNA
;   ORGANISM: Artificial sequence
;   FEATURE:
;   NAME/KEY: promoter
;   LOCATION: (1)..(1141)
;   OTHER INFORMATION: consensus sequence of A.t., L.a., and B.n. FAE1 promoters
US-09-806-708B-22

Query Match   2.5%; Score 47.6; DB 4; Length 1141;
Best Local Similarity   9.5%; Pred. No. 0.0016;
Matches   36; Conservative 172; Mismatches 169; Indels   0; Gaps   0;

```
Qy    1145 TTTCTTCCGATTCGTTGAGAAATTAAAGTCAAATGATGAATTGAAAAGTGCATTGGACA 1204
                :::  :::::::  |::::: ::: :: |::: |::| :|:::: ::|:| ||:| ::
Db      83 YKKKWYBCANNTSBRYHARRWKDMKTAYBMTWTNKWGKTGWRHRYWRWRAMBDTVDHHYV 142

Qy    1205 GAATTAAATATCCGAGTTGATGACGTGGAAAAGTATCTTGCTCCAAAGCCGATGGAATTCA 1264
                ||:: |:: |::::::::: ||: |:::::: :::: |: :::: |:: :::|:: :
Db     143 TAMNNAWTTMCMMKDDKDKRTRHWWKKNNNATGWDDDTKYHMWNNNGCBTVTWMVRYKTDR 202

Qy    1265 ATCCAAAACCTCAGCCTAGCGCTGGCTACTTTGCTCCACGTAAAATCCCAACAAGACCACGTAAAA 1324
                ::::: |: :::::::::::::::: ::::: :  :||::: |:: :: : :::::::
Db     203 DWSBKRMNYGMBWWKNWSYDVTYYWWVWDDMCKRKVRRWVRTRGRMRNYMVAWBTAHRRR 262
```

RESULT 6
US-09-007-005-17
; Sequence 17, Application US/09007005B
; Patent No. 6258558
; GENERAL INFORMATION:
; APPLICANT: Szostak, Jack W.
; APPLICANT: Roberts, Richard W.
; APPLICANT: Liu, Rihe

```
Qy    1325 TGCTTCCATTATTAATTGGATCTGATCCAAAAGTTCAAGAGGAAATACGAAGACATCCAA 1384
                ||:|        :|:   :       ::|:| :|::|::|:|:::|:::::::::|:
Db     263 YNNGWTBAMAYRRWTMNNNNNAKAMCKRAKYWGWNRABVNSTCTTWKSKTTKVRTSCWA 322

Qy    1385 GTACCGAATGGAAAATTGCAAAAGAATCAAGAGAGTTTGACAAATTTGAAGAATAATCCAA 1444
                |::| ::: :::::: ::: :: :::::::| |::::::::::::| :: ::::::| :: ::
Db     323 NNCRAGDANKDHKWKWSAAMGVYWNNNMNNNWTYKKARHBARWDWVWHSAWKKWHANAA 382

Qy    1445 GTCTTGCTGCATTGTTCATGGATGATGAATAAATTAGAGAATACATTGAAAGGAAGGCAAATGT 1504
                :: :|::| :::|::::: :: :::: ::::::  ::|:|: |:::: :::::::|
Db     383 HYSRKKWTBYKRKTMVNNNGTTMWKRMWAWYWKMDMDWBGTYNNNNNGGRTYYGWTKNK 442

Qy    1505 TAACTGATGAACAGAAA 1521
                :: :| ::   ::|
Db     443 KMWTYYKWKANNCKWRA 459
```

RESULT 5
US-09-248-796A-6002
; Sequence 6002, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstock et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796A
; CURRENT FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
;
; SEQ ID NO 6002
;   LENGTH: 1086
;   TYPE: DNA
;   ORGANISM: Candida albicans
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (987),(988)
;   OTHER INFORMATION: Identity of nucleotide sequences at the above locations are unkno
US-09-248-796A-6002

Query Match   46.8; DB 4; Length 1086;
Best Local Similarity 50.9%; Pred. No. 0.0028;
Matches 111; Conservative   0; Mismatches 107; Indels   0; Gaps   0;

```
Qy     131 TACCACTGGCTCCATCAATGGAAGCTCTTGAACTGATGGGTGTTCAATTTGTTGATGCTC 190
               ||||||||||||||||||||||||||||| ||||||||||||| ||||||| | |||
Db     626 TACCAATGGCTTTAGCAAGTGGTGCTGCTGAAATGGCAAAAGGAGCATTTAATGGGTATTT 685

Qy     191 TCATCAAAAAGTCAAATGGAAATAACGGAAGGAGAAATTCTATTGAAGCAACTCCAGAATGA 250
               ||||| ||||| |||||| ||  ||  ||| ||:|||:| ||||||||:|| |||||||||
Db     686 TCATGTCTCAAGGTGAAATAACGCGCAAGGAGAAATTCTATTGAAGCAACTCCAGAAATGA 745

Qy     251 TAGAGAAAGTACATCCTGATCAATTCGATAAGTAAGTACAAAAAGTGATGATTCG 310
               ||||||  ||||| ||||||| ||||||||  |||||||||  |||||||||||
Db     746 AACAATTAATGAACAACAGATGATTTAGCTATAGAAGTAGATCATGCTGATTTTGAATGGG 805

Qy     311 CAGCTGATGCAGTTATGCAACAGCGGAGATGGCAAAA 348
               ||||||||| |||||||||  |||||||||| ||||
Db     806 AAACTTTTGCCGATGAAGAAGATCCAGATAAGAAAAAAA 843
```

; TITLE OF INVENTION: SELECTION OF PROTEINS USING RNA-PROTEIN
; TITLE OF INVENTION: FUSIONS
; FILE REFERENCE: 00786/350003
; CURRENT APPLICATION NUMBER: US/09/007,005B
; CURRENT FILING DATE: 1998-01-14
; EARLIER APPLICATION NUMBER: 60/035,963
; EARLIER FILING DATE: 1997-01-27
; EARLIER APPLICATION NUMBER: 60/064,491
; EARLIER FILING DATE: 1997-11-06
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 17
; LENGTH: 289
; TYPE: RNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Translation template
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(289)
; OTHER INFORMATION: n = A,T,C or G
US-09-007-005-17

Query Match   2.4%;   Score 46;   DB 3;   Length 289;
Best Local Similarity 10.0%;   Pred. No. 0.002;
Matches   22; Conservative 93; Mismatches 105; Indels   0; Gaps   0;

Qy   621 ACAAGCTCTTGCTCAAAGAGCAATGTTAGGTAAAAATGCCCCAGTTGCCGGTGGAAGAGG 680
              ::: ::: : : :::::::: :   :: ::: :: ::: : : : :: : :::
Db   67 RSRNRNRSRNRNRSRNRNRSRNRNRSRNRNRSRNRNRSRNRNRSRNRNRSRNRNRN 126

Qy   681 TGAAGAACAACGGATGATGATGAATCGAGTGACCAAAGAATGCAACAAAGAGAACTTCA 740
              :: ::: ::   ::: ::: ::: :::: : ::: ::
Db   127 RSRNRNRSRNRNRSRNRNRSRNRNRSRNRNRSRNRNRSRNRNRSRNRNRSRNRNRN 186

Qy   741 AGAGGAAGATGAAGATGATGATCTTGAGGACGAGGATGTACCCAGAAGAAGAAGTTC 800
              ::: ::: ::: :: : :::: :|::|:|::| |:|:: |:|:: |:::|
Db   187 RSRNRNRSRNRNRSRNRNRSRNRNRSRCRARGRCRGRURARARCRURC 246

Qy   801 GGATGGAGAACCACAAAGTGAAGCAGAGCATCAGAGAAGA 840
              : ::::| : :||| |||| |||||||||||
Db   247 RURURGRCRGRCRUAAAAAAAAAAAAAAAAAAAA 286

RESULT 7
US-09-244-796-17
; Sequence 17, Application US/09244796
; Patent No. 6281344
; GENERAL INFORMATION:
; APPLICANT: Szostak, Jack W.
; APPLICANT: Roberts, Richard W.
; APPLICANT: Liu, Rihe
; TITLE OF INVENTION: SELECTION OF PROTEINS USING RNA-PROTEIN
; TITLE OF INVENTION: FUSIONS
; FILE REFERENCE: 00786/350007
; CURRENT APPLICATION NUMBER: US/09/244,796
; CURRENT FILING DATE: 1999-02-05
; EARLIER APPLICATION NUMBER: 60/035,963
; EARLIER FILING DATE: 1997-01-27
; EARLIER APPLICATION NUMBER: 60/064,491
; EARLIER FILING DATE: 1997-11-06
; EARLIER APPLICATION NUMBER: 09/007,005
; EARLIER FILING DATE: 1998-01-14
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 17
; LENGTH: 289
; TYPE: RNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Translation template
; FEATURE:
; NAME/KEY: misc_feature

; LOCATION: (1)..(289)
; OTHER INFORMATION: n = A,T,C or G
US-09-244-796-17

Query Match   2.4%;   Score 46;   DB 3;   Length 289;
Best Local Similarity 10.0%;   Pred. No. 0.002;
Matches   22; Conservative 93; Mismatches 105; Indels   0; Gaps   0;

Qy   621 ACAAGCTCTTGCTCAAAGAGCAATGTTAGGTAAAAATGCCCCAGTTGCCGGTGGAAGAGG 680
              ::: ::: : : :::::::: :   :: ::: :: ::: : : : :: : :::
Db   67 RSRNRNRSRNRNRSRNRNRSRNRNRSRNRNRSRNRNRSRNRNRSRNRNRSRNRNRN 126

Qy   681 TGAAGAACAACGGATGATGATGAATCGAGTGACCAAAGAATGCAACAAAGAGAACTTCA 740
              :: ::: ::   ::: ::: ::: :::: : ::: ::
Db   127 RSRNRNRSRNRNRSRNRNRSRNRNRSRNRNRSRNRNRSRNRNRSRNRNRSRNRNRN 186

Qy   741 AGAGGAAGATGAAGATGATGATCTTGAGGACGAGGATGTACCCAGAAGAAGAAGTTC 800
              ::: ::: ::: :: : :::: :|::|:|::| |:|:: |:|:: |:::|
Db   187 RSRNRNRSRNRNRSRNRNRSRNRNRSRCRARGRCRGRURARARCRURC 246

Qy   801 GGATGGAGAACCACAAAGTGAAGCAGAGCATCAGAGAAGA 840
              ::::::| |: :|||| |||
Db   247 RURURGRCRGRCRUAAAAAAAAAAAAAAAAAAAA 286

RESULT 8
US-09-949-016-14193/c
; Sequence 14193, Application US/0994916
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 14193
; LENGTH: 247781
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(247781)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14193

Query Match   2.4%;   Score 45.8; DB 4;   Length 247781;
Best Local Similarity 57.2%;   Pred. No. 0.22;
Matches   83; Conservative   0; Mismatches 62; Indels   0; Gaps   0;

Qy   682 GAAGAACAACGGATGATGATGAATCGAGTGACCAAAGAATGCAACAAAGAGAACTTCAA 741
              ||||| ||||||||||| ||| ||| |||||||||||||||||||||||||||
Db 126717 GAAGAGGACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGAG 126658

Qy   742 GAGGAAGATGAAGATGATGATGATCTTGAGGACGAGGATGTACCCAGAAGAAGAAGTTCG 801
              |||||||||||||||||||| |||||||||||||||||||||||||||||||||
Db 126657 GAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAAGAGAAGAAGAAGA 126598

Qy   802 GATGGAGAACCACAAAGTGAAGCAG 826
              |||| |||| ||||||| |||||
Db 126597 GAAGGAGGAGGAGGAGGAGGAAGACAG 126573

RESULT 9
US-09-949-016-86857/c

```
; Sequence 86857, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;   APPLICANT: VENTER, J. Craig et al.
;   TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;   TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;   FILE REFERENCE: CL001307
;   CURRENT APPLICATION NUMBER: US/09/949,016
;   CURRENT FILING DATE:  2000-04-14
;   PRIOR APPLICATION NUMBER: 60/241,755
;   PRIOR FILING DATE: 2000-10-20
;   PRIOR APPLICATION NUMBER: 60/237,768
;   PRIOR FILING DATE: 2000-10-03
;   PRIOR APPLICATION NUMBER: 60/231,498
;   PRIOR FILING DATE: 2000-09-08
;   NUMBER OF SEQ ID NOS: 207012
;   SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 86857
;   LENGTH: 601
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-86857

Query Match            2.3%;  Score 44.2;  DB 4;  Length 601;
Best Local Similarity 58.9%;  Pred. No. 0.011;
Matches   76;  Conservative    0;  Mismatches   53;  Indels    0;  Gaps    0;

Qy      682 GAAGAACAACGATGATGATGAATCGAGTCGACCAAAGAAATGCAACAAAGAGAACTTCAA 741
            |||||||   ||| |||  ||| |||   |||| | || |||||||||||| ||| |||
Db      132 GAAGAGGACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 73

Qy      742 GAGGAAGATGAAGATGAAGATGATGATGAAGATCTTGAGGACGAGGATGTACCCAGAAGAAGTTCG 801
            ||||||| || |||| ||| ||| | ||||||||||||||||||||| ||||||||
Db       72 GAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAAGA 13

Qy      802 GATGGAGAA 810
            || |||||
Db       12 GAAGGAGGA 4


RESULT 10
US-09-949-016-14581/c
; Sequence 14581, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;   APPLICANT: VENTER, J. Craig et al.
;   TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;   TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;   FILE REFERENCE: CL001307
;   CURRENT APPLICATION NUMBER: US/09/949,016
;   CURRENT FILING DATE:  2000-04-14
;   PRIOR APPLICATION NUMBER: 60/241,755
;   PRIOR FILING DATE: 2000-10-20
;   PRIOR APPLICATION NUMBER: 60/237,768
;   PRIOR FILING DATE: 2000-10-03
;   PRIOR APPLICATION NUMBER: 60/231,498
;   PRIOR FILING DATE: 2000-09-08
;   NUMBER OF SEQ ID NOS: 207012
;   SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 14581
;   LENGTH: 64309
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;     NAME/KEY: misc_feature
;     LOCATION: (1)..(64309)
;     OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14581

Query Match            2.3%;  Score 43.8;  DB 4;  Length 64309;
Best Local Similarity 50.2%;  Pred. No. 0.35;
Matches  108;  Conservative    0;  Mismatches  107;  Indels    0;  Gaps    0;
```

```
Qy      628 CTTGCTCAAAGAGCAATGTTAGTGTAAAAATGCCCAGTTGCCGGTGGAAGAGGTGAAGAA 687
            | |||  |    | |||||           ||   |    |||   | | ||| |||||
Db    24093 CTGGGTGACAGAGGCAAACTCCATGAGAGAAGATGAAGGAGGAAGGAACTTCAAGAGGAA 24034

Qy      688 CAACGGATGATGATGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGAAGAACTTCAAGAGGAA 747
            |||                |||||||||||||||||||||| |||||||||||| |||||
Db    24033 GAAGAAGAGGAGGAGGAGGAGGAGGAGGAGGATGTACCCAGAAGAAGAGAAGAGAAGAAGAA 23974

Qy      748 GATGAAGATGATGATGATCTTGAGGACGAGGATGTACCCAGAAGAAGAGAAGTTCGGATGGA 807
            || ||||||| |||| || |||||||||||||||||||||| ||||
Db    23973 GAAGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGCATCAGAGAAGGA 23914

Qy      808 GAACCACAAAGTGAAGCAGCAGAGCATCAGAGAAGAGA 842
            |||  || ||   |||||  |||| ||| ||||||||
Db    23913 GAAGAGAAGAGGAGGAGGAGGAGGAGAAGAGAGGAAGA 23879


RESULT 11
US-09-949-016-91474
; Sequence 91474, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;   APPLICANT: VENTER, J. Craig et al.
;   TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;   TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;   FILE REFERENCE: CL001307
;   CURRENT APPLICATION NUMBER: US/09/949,016
;   CURRENT FILING DATE:  2000-04-14
;   PRIOR APPLICATION NUMBER: 60/241,755
;   PRIOR FILING DATE: 2000-10-20
;   PRIOR APPLICATION NUMBER: 60/237,768
;   PRIOR FILING DATE: 2000-10-03
;   PRIOR APPLICATION NUMBER: 60/231,498
;   PRIOR FILING DATE: 2000-09-08
;   NUMBER OF SEQ ID NOS: 207012
;   SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 91474
;   LENGTH: 601
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-91474

Query Match            2.3%;  Score 43.6;  DB 4;  Length 601;
Best Local Similarity 53.5%;  Pred. No. 0.017;
Matches   91;  Conservative    0;  Mismatches   79;  Indels    0;  Gaps    0;

Qy      422 GTGGATTAGAAGATGCTATCCGAACGCAGAGAGATATGGAAAATACGGATCCGTCCGAAC 481
            |||||  ||| |||  || ||||| || |||||||||  |||||| |   | | |||||
Db      142 GGGAATCAGTTAACCCAGCCCCACTCAAGGGGATAGGGACTGGGCTCCACCTTTTGAAG 201

Qy      482 AGATTGCCAAAGCCGTAATGGCGTAATGAC ATTTCAAACACAAATTCTCCCAGGACTCGTTGCAA 541
            ||||| ||||| |||||||| |||||||||||||||  ||||||||||||||
Db      202 AGAATATCAAAGAATTTGTGGACATATTTAAACTACCATGCTATGTTTAATTATGGCAA 261

Qy      542 ATATGATCGTGGCAAGAACCCCTTTAAAATGCCTCAACAAATGAGAAAA 591
            || |||||||||| ||||| ||||| ||| |  |||  ||| ||||||
Db      262 GTATGGTCACTGTAGAAAACTGGAAAATAGGACAACGCAACAGCAAAAAAAAA 311


RESULT 12
US-09-949-016-91475
; Sequence 91475, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;   APPLICANT: VENTER, J. Craig et al.
;   TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;   TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;   FILE REFERENCE: CL001307
;   CURRENT APPLICATION NUMBER: US/09/949,016
;   CURRENT FILING DATE:  2000-04-14
;   PRIOR APPLICATION NUMBER: 60/241,755
;   PRIOR FILING DATE: 2000-10-20
;   PRIOR APPLICATION NUMBER: 60/237,768
```

RESULT 14
US-09-248-796A-3448
; Sequence 3448, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstock et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICANS
; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796A
; CURRENT FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 3448
; LENGTH: 1755
; TYPE: DNA
; ORGANISM: Candida albicans
US-09-248-796A-3448

Query Match   2.2%;   Score 43.2; DB 4;   Length 1755;
Best Local Similarity   58.6%;   Pred. No. 0.047;
Matches   75;   Conservative   0;   Mismatches   53;   Indels   0;   Gaps   0;

Qy   692 GGATGATGATGAGTGAGTGGACCAACAAGAATGCAACAAGAGAACTTCAAGAGGAAGATG 751
                |||||||||||||| || | ||| |||||||||||||| ||| || ||||||||
Db   1625 GGAGGAAGAGTAATAATAATAATCAAATAATCAAATAATTAAAGAAGAAGGATTGAACATAATGATG 1684

Qy   752 AAGATGATGATGATCTTGAGGACGAGGATGTACCCAGAAGAAGAAGAGTTCGGATGGAGAAC 811
                | |||||||||||||| ||| ||| |||||||||| ||||||||||||||||||||||
Db   1685 ACGATGATGATGATGAAGATGAAGTGAAGTAAGTGAAAACGAGAGATTCCAACTTATGGATATC 1744

Qy   812 CACAAAGT 819
                ||||||||
Db   1745 AACAGAAT 1752

RESULT 15
US-09-820-007-3/c
; Sequence 3, Application US/09820007
; Patent No. 6830900
; GENERAL INFORMATION:
; APPLICANT: YAN, Chunhua et al
; TITLE OF INVENTION: ISOLATED HUMAN TRANSPORTER PROTEINS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN TRANSPORTER PROTEINS,
; TITLE OF INVENTION: AND USES THEREOF
; FILE REFERENCE: CL001205
; CURRENT APPLICATION NUMBER: US/09/820,007
; CURRENT FILING DATE: 2001-03-29
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 213456;
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(213456)
; OTHER INFORMATION: n = A,T,C or G
US-09-820-007-3

Query Match   2.2%;   Score 42.2; DB 4;   Length 213456;
Best Local Similarity   51.9%;   Pred. No. 2.4;
Matches   95;   Conservative   0;   Mismatches   88;   Indels   0;   Gaps   0;

Qy   1299 TAAAATCCAACAAGACCACGTAAAATGCTTCCATTATTATTAATTGGATCTGATCATCAAAAGT 1358
                || ||| || ||| |||| ||| | ||| | ||| | | ||| ||| | |||||||
Db 102305 TTAAAACTAACAAATGCAGGTTTACTGCTTTTATGATAACTTGGTATAGATAACAGTT 102246

Qy   1359 TCAAGACGAAATACGAAGACATCCAAGTACCCAGATGAAAATTGCAAAAGAATCAAGAGT 1418

RESULT 13
US-09-949-016-14370
; Sequence 14370, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 14370
; LENGTH: 107980
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(107980)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14370

Query Match   2.3%;   Score 43.6; DB 4;   Length 107980;
Best Local Similarity   53.5%;   Pred. No. 0.57;
Matches   91;   Conservative   0;   Mismatches   79;   Indels   0;   Gaps   0;

Qy   422 GTGGATTAGAAGATGCTATCCGAACCGAGAGAGATATGGAAAATACGGATCCGTCCGAAC 481
                 | | | || || || | ||||||| |||||||||| ||| |||| || |||||
Db   7404 GGGAATCAGTTAACCCAGCCCCACTCAAGGGGATAGGGACTGGGCTCCACCTTTTGAAG 7463

Qy   482 AGATTGCCAAAGCCGTAATGGACAAATTCAAACACAAATTCTCCAGGACTCGTTGCAA 541
                |||||||||| ||| ||||||| |||||||||||| |||||||||||||| ||||
Db   7464 AGAATATCAAAGAATTTGTGGACATATTTAAACTACCATGCTATGTTTAATTATGGCAA 7523

Qy   542 ATATGATCGCTGGCAAGAACCCCTTTAAAATGCCTCAACAAATGAGAAAA 591
                ||||||||| ||||||| | || |||| | | ||| ||| || |||||
Db   7524 GTATGGTCACTGTAGAAAAACTGGAAAAATAGGGACAAGCAAGAAAAAAAAAA 7573

Query Match   2.3%;   Score 43.6; DB 4;   Length 601;
Best Local Similarity   53.5%;   Pred. No. 0.017;
Matches   91;   Conservative   0;   Mismatches   79;   Indels   0;   Gaps   0;

Qy   422 GTGGATTAGAAGATGCTATCCGAACGCAGAGAGATATGGAAAATACGGATCCGTCCGAAC 481
                 | | | || || || | ||||||| |||||||||| ||| |||| || |||||
Db   96 GGGAATCAGTTAACCCAGCCCCACTCAAGGGGATAGGGACTGGGCTCCACCTTTTGAAG 155

Qy   482 AGATTGCCAAAGCCGTAATGGACAAATTCAAACACAAATTCTCCAGGACTCGTTGCAA 541
                ||||||||||| ||| ||||||| |||||||||||| |||||||||||||| ||||
Db   156 AGAATATCAAAGAATTTGTGGACATATTTTAAACTACCATGCTATGTTTAATTATGGCAA 215

Qy   542 ATATGATCGCTGGCAAGAACCCCTTTAAAATGCCTCAACAAATGAGAAAA 591
                ||||||||| ||||||| | || |||| | | ||| ||| || |||||
Db   216 GTATGGTCACTGTAGAAAAACTGGAAAAATAGGGACAAGCAAGCAAAAAAAAAAA 265

; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 91475
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-91475

```
Db     102245 TGAAGAGGTAAAATAAGGATGCCCTTGTTTCACAAATTTTTTTTTAGGAAAATATAGGTG 102186
              |||||||| ||  | ||  ||  ||  |    |   |  |  | |||| ||   ||
Qy       1419 TTTGACAAATTTGAAGAATAATCCAAGTCTTGCTGCATTGTTCATGGATGATAAATTAGA 1478
              |||||||       |||| |  |       |  |  | | || ||||||||
Db     102185 TAAAGTAAATTTCTTTAAAATTGTATCATTTCCACAAAGCCTGTTATTGGTAAATTAGA 102126

Qy       1479 GAA 1481
              |||
Db     102125 GAA 102123
```

Search completed: July 15, 2005, 15:21:19
Job time : 347 secs

```
                    GenCore version 5.1.6
              Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on:        July 15, 2005, 10:47:26 ; Search time 1219 Seconds
                                         (without alignments)
                              9972.607 Million cell updates/sec

Title:         US-10-736-868-1
Perfect score: 1932
Sequence:      1 atgattttatttcattttt........tgcctgcactgggattgtag 1932

Scoring table: IDENTITY_NUC
               Gapop 10.0 , Gapext 1.0

Searched:      6340675 seqs, 3146120178 residues

Total number of hits satisfying chosen parameters:    12681350

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :     Published Applications NA:*
          1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
          2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
          3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
          4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
          5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
          6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
          7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
          8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
          9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:*
         10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*
         11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq:*
         12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
         13: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
         14: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
         15: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq:*
         16: /cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq:*
         17: /cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq:*
         18: /cgn2_6/ptodata/1/pubpna/US10F_PUBCOMB.seq:*
         19: /cgn2_6/ptodata/1/pubpna/US10G_PUBCOMB.seq:*
         20: /cgn2_6/ptodata/1/pubpna/US10H_PUBCOMB.seq:*
         21: /cgn2_6/ptodata/1/pubpna/US10I_PUBCOMB.seq:*
         22: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
         23: /cgn2_6/ptodata/1/pubpna/US11A_PUBCOMB.seq:*
         24: /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq:*
         25: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
         26: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1932	100.0	1932	21	US-10-736-868-1	Sequence 1, Appli
c 2	50.4	2.6	193853	13	US-10-087-192-1663	Sequence 1663, Ap
c 3	47.2	2.4	3534	17	US-10-282-122A-16250	Sequence 16250, A
c 4	46.2	2.4	96602	17	US-10-085-117-61	Sequence 61, Appl
c 5	45.8	2.4	17848	14	US-10-239-676-28	Sequence 28, Appl
c 6	45.8	2.4	17848	15	US-10-240-453-38	Sequence 38, Appl
c 7	45.8	2.4	17848	17	US-10-257-166-58	Sequence 58, Appl

```
   8   45.8  2.4  127767  19  US-10-367-094-176          Sequence 176, App
   9   45.6  2.4     548  19  US-09-864-761-18908        Sequence 18908, A
c 10   45.4  2.3     272   9  US-09-864-761-18684        Sequence 18684, A
c 11   45.4  2.3  153170  20  US-10-723-860-199          Sequence 199, App
c 12   45    2.3  154817  17  US-10-085-117-334          Sequence 334, App
  13   44.8  2.3     543  16  US-10-029-386-3612         Sequence 3612, Ap
  14   44.6  2.3     694  18  US-10-424-599-83149        Sequence 83149, A
  15   44.6  2.3     694  18  US-10-425-114-11335        Sequence 11335, A
c 16   44.6  2.3   61876  13  US-10-087-192-1327         Sequence 1327, Ap
c 17   44.4  2.3     392  19  US-10-674-124A-155         Sequence 155, Appl
c 18   44.4  2.3   32069  16  US-10-004-113-7            Sequence 7, Appli
  19   44.4  2.3  224112  19  US-10-367-094-80           Sequence 80, Appl
  20   44.2  2.3     403  19  US-10-674-124A-22585       Sequence 22585, A
c 21   44.2  2.3     492  19  US-10-674-124A-21270       Sequence 21270, A
c 22   44.2  2.3   61791  19  US-10-322-281-645          Sequence 645, App
c 23   44.2  2.3  684187  19  US-10-367-094-71           Sequence 71, Appl
  24   43.6  2.3    1723  17  US-10-104-047-1498         Sequence 1498, Ap
  25   43.4  2.2     457  19  US-10-674-124A-5649        Sequence 5649, Ap
c 26   43.4  2.2  219715  20  US-10-417-375-63           Sequence 63, Appl
c 27   43.2  2.2     848  13  US-10-027-632-144785       Sequence 144785,
c 28   43.2  2.2     848  17  US-10-027-632-144785       Sequence 144785,
  29   43.2  2.2    1752  19  US-10-451-467A-411         Sequence 411, App
  30   43    2.2     387  19  US-10-674-124A-19168       Sequence 19168, A
  31   43    2.2     435  10  US-09-814-353-5103         Sequence 5103, Ap
c 32   43    2.2     435  10  US-09-814-353-11395        Sequence 11395, A
c 33   43    2.2     447  19  US-10-674-124A-6601        Sequence 6601, Ap
  34   43    2.2     536  16  US-10-029-386-19974        Sequence 19974, A
  35   43    2.2     599  16  US-10-029-386-6243         Sequence 6243, Ap
  36   43    2.2    1702  18  US-10-114-270-177          Sequence 177, App
  37   43    2.2  312477  19  US-10-317-883A-12          Sequence 12, Appli
c 38   43    2.2  1691139 14  US-10-067-514-1            Sequence 1, Appli
c 39   43    2.2  1691139 17  US-10-419-723-1            Sequence 1, Appli
  40   42.8  2.2     223  16  US-10-029-386-17312        Sequence 17312, A
  41   42.8  2.2    7168  18  US-10-240-589C-39          Sequence 39, Appl
c 42   42.8  2.2  421609  19  US-10-367-094-122          Sequence 122, App
c 43   42.6  2.2     245  19  US-10-674-124A-16652       Sequence 16652, A
  44   42.6  2.2     759  20  US-10-425-115-157981       Sequence 157981,
c 45   42.6  2.2   82121  17  US-10-085-117-136          Sequence 136, App
```

ALIGNMENTS

RESULT 1
US-10-736-868-1
; Sequence 1, Application US/10736868
; Publication No. US20050079160A1
; GENERAL INFORMATION:
; APPLICANT: Solomon, Aaron
; APPLICANT: Morimoto, Richard
; APPLICANT: Beitel, Greg
; TITLE OF INVENTION: OSR-1 Nucleic Acids and Proteins
; FILE REFERENCE: NWESTERN-08451
; CURRENT APPLICATION NUMBER: US/10/736,868
; CURRENT FILING DATE: 2003-12-16
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1
; LENGTH: 1932
; TYPE: DNA
; ORGANISM: Caenorhabditis elegans
US-10-736-868-1

Query Match   100.0%; Score 1932; DB 21; Length 1932;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1932; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATGATTTTATTTCATTTTTATTTTTATTTTTATTCCTGTGTTGTAGGATTTGTATCGCACCATTATCGGCC 60
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1 ATGATTTTATTTCATTTTTATTTTTATTTTTATTCCTGTGTTGTAGGATTTGTATCGCACCATTATCGGCC 60

Qy 61 CAATCTCCATCGACTTCCGATGCTTCCGGGAGCTTGTTGTTGTCATCTCTCGTAGTAAAAGC 120
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

```
Db     61  CAATCTCCATCGACTTCCGATGCTCCGGGAGCTTGTTGTCATCTCTCGTAGGTAAAAGC 120

Qy    121  CATCAAAAACTACCACTGGCTCCATCAATGGAAGCTCTTGAACTGATGGGTGTTCAATTT 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121  CATCAAAAACTACCACTGGCTCCATCAATGGAAGCTCTTGAACTGATGGGTGTTCAATTT 180

Qy    181  GTTGATGCTCTCATCAAAAAGGTCAAAATGCAAAAGGAGCATTTAAGACTCAA 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181  GTTGATGCTCTCATCAAAAAGGTCAAAATGCAAAAGGAGCATTTAAGACTCAA 240

Qy    241  TTAGAGAGTTCTAGAGAAATACATCCTGATCAATTCGATAAGTAGTACAAAAAGCTAAAAGTT 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241  TTAGAGAGTTCTAGAGAAATACATCCTGATCAATTCGATAAGTAGTACAAAAAGCTAAAAGTT 300

Qy    301  GATGATTTGGCAGCTGATGCAGTTGCAACAGGCGGAGATGGCAAAAATTACAGCCTAAA 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301  GATGATTTGGCAGCTGATGCAGTTGCAACAGGCGGAGATGGCAAAAATTACAGCCTAAA 360

Qy    361  TCAGGAAATGCATTTATCGATATGTTGAATGGAATCCCAATTGGTAGCAGTATT 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361  TCAGGAAATGCATTTATCGATATGTTGAATGGAATCCCAATTGGTAGCAGTATT 420

Qy    421  CGTGGATTAGAAGATGCTATCCGAACGCAGAGATATGGAAAATACCGATCCGTCCGAA 480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421  CGTGGATTAGAAGATGCTATCCGAACGCAGAGATATGGAAAATACCGATCCGTCCGAA 480

Qy    481  CAGATTGCCAAAGCCGTAATGCAAAATTCAAACAACAAATTCTCCCAGACTCGTTGCA 540
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481  CAGATTGCCAAAGCCGTAATGCAAAATTCAAACAACAAATTCTCCCAGACTCGTTGCA 540

Qy    541  AATATGATCGCTGGCAAGAACCCCTTTAAAATGCTCAACAAATGAGAAAGTCAAGCT 600
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541  AATATGATCGCTGGCAAGAACCCCTTTAAAATGCTCAACAAATGAGAAAGTCAAGCT 600

Qy    601  GCTCCATCGTCAGTGTTCCAACAAGCTCTTGCTCAAAGAGCAATGTTAGTGTAAAATGCC 660
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601  GCTCCATCGTCAGTGTTCCAACAAGCTCTTGCTCAAAGAGCAATGTTAGTGTAAAATGCC 660

Qy    661  CCAGTTGCCGGTGAAGAGAGGTGAAGAACAACGGATGATGATGATGAGTGGACCAAAGA 720
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661  CCAGTTGCCGGTGAAGAGAGGTGAAGAACAACGGATGATGATGATGAGTGGACCAAAGA 720

Qy    721  ATGCAACAACAGAAGAACTTCAAGAGAAGATGAAGATGATGATGAGGACGAGGAT 780
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    721  ATGCAACAACAGAAGAACTTCAAGAGAAGATGAAGATGATGATGAGGACGAGGAT 780

Qy    781  GTACCCAGAAGCAGAAGAGTTCGGATGGAGAACCACAAAGTGAAGCAGAGCATCAGAGAAGA 840
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    781  GTACCCAGAAGCAGAAGAGTTCGGATGGAGAACCACAAAGTGAAGCAGAGCATCAGAGAAGA 840

Qy    841  GATTTAGCCAGGAGCAGATTGAAAAGAAGTAGTCCTAGATTAAAAGAGAGCTTTTACAGAATGCGGAA 900
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    841  GATTTAGCCAGGAGCAGATTGAAAAGAAGTAGTCCTAGATTAAAAGAGAGCTTTTACAGAATGCGGAA 900

Qy    901  GTTCAATCATTGCTCTCTTACCAACGAATGAGGGATTCTCCACTGAGCAAGCGAAGCCT 960
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    901  GTTCAATCATTGCTCTCTTACCAACGAATGAGGGATTCTCCACTGAGCAAGCGAAGCCT 960

Qy    961  TTGGCTATGAACGATGAGGATGAAAGTGCATTCCGCGAATGGAGGCTCGTGCAAAACTA 1020
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    961  TTGGCTATGAACGATGAGGATGAAAGTGCATTCCGCGAATGGAGGCTCGTGCAAAACTA 1020

Qy   1021  GATCAAAATCTCAACTTGTGCTCGGTCTCGTTCTCCATGGTTTTGGAGAGTCTGATGATGATGAA 1080
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1021  GATCAAAATCTCAACTTGTGCTCGGTCTCGTTCTCCATGGTTTTGGAGAGTCTGATGATGATGAA 1080

Qy   1081  GACGAAGAAGATGAAAATTGAAAATTTGATTGATCCATCTGAAAATTCATTCCGTCGTGCACCACTT 1140
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1081  GACGAAGAAGATGAAAATTGAAAATTTGATTGATCCATCTGAAAATTCATTCCGTCGTGCACCACTT 1140

Qy   1141  CGTCTTTTCTTCCGGATTCGTTGAGAAATTAAAGTCAAATGAATTGAATATGATGAAAGTGCATTG 1200
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1141  CGTCTTTTCTTCCGGATTCGTTGAGAAATTAAAGTCAAATGAATTGAATATGATGAAAGTGCATTG 1200
```

```
Qy   1201  GACAGAATTAAATATGAGTTGATGACGTGAAAAGTATCTTGCTCCAAAGCCGATGCGAA 1260
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1201  GACAGAATTAAATATGAGTTGATGACGTGAAAAGTATCTTGCTCCAAAGCCGATGCGAA 1260

Qy   1261  TTCAATCCAAAAACCTCAGCCTGGCTACTTGCTCCACGTAAAATCCAACAGCAAGAGACCACGT 1320
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1261  TTCAATCCAAAAACCTCAGCCTGGCTACTTGCTCCACGTAAAATCCAACAGCAAGAGACCACGT 1320

Qy   1321  AAAATGTTCCATTATTAATTAATTGGATCTGATCCAAAAGTTCAAGAGGAAATACGAAGACAT 1380
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1321  AAAATGTTCCATTATTAATTAATTGGATCTGATCCAAAAGTTCAAGAGGAAATACGAAGACAT 1380

Qy   1381  CCAAGTACCGAATGGAAAATTGCAAAAGAATCAAGAGTTTGACAAATTTGAAGAATAAT 1440
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1381  CCAAGTACCGAATGGAAAATTGCAAAAGAATCAAGAGTTTGACAAATTTGAAGAATAAT 1440

Qy   1441  CCAAGTCTTGCTGCATTGTGCATTGTTCAAACACGTGTCAAAACAATTCGTGCATTACCAAGA 1500
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1441  CCAAGTCTTGCTGCATTGTGCATTGTTCAAACACGTGTCAAAACAATTCGTGCATTACCAAGA 1500

Qy   1501  ATGTTAACTGACAACAGAAAGGTAGAACACGTGTCAAAATGATGATGAAGTGTATTCCAAGATATT 1560
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1501  ATGTTAACTGACAACAGAAAGGTAGAACACGTGTCAAAATGATGATGAAGTGTATTCCAAGATATT 1560

Qy   1561  CTGTTCGGTGACCAACTGCAAAAGTGAAATGATTGATGCAAAAGGTATTCCCATCT 1620
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1561  CTGTTCGGTGACCAACTGCAAAAGTGAAATGATTGATGCAAAAGGTATTCCCATCT 1620

Qy   1621  GAAGAACGTCCCATTCCTCCATTGTTCTTGAACCAAAGAAGGAAGGCATACGAGATTGAGA 1680
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1621  GAAGAACGTCCCATTCCTCCATTGTTCTTGAACCAAAGAAGGAAGGCATACGAGATTGAGA 1680

Qy   1681  TGGACTGGAGCAAATGAAAAGAAATTCCAGACTTGGAAGTCGCTTCATTCTCCCATCT 1740
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1681  TGGACTGGAGCAAATGAAAAGAAATTCCAGACTTGGAAGTCGCTTCATTCTCCCATCT 1740

Qy   1741  CTTGATCCAACTATGCCAGCCTTGAACACGGCTTTCTCGACTCAGGGGCGAGCCCGTGAC 1800
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1741  CTTGATCCAACTATGCCAGCCTTGAACACGGCTTTCTCGACTCAGGGGCGAGCCCGTGAC 1800

Qy   1801  GAGTGGGATACCATGTTCAAAATCCCGAATAACTGGAATTCTGGAGATGAAGTGGGTTC 1860
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1801  GAGTGGGATACCATGTTCAAAATCCCGAATAACTGGAATTCTGGAGATGAAGTGGGTTC 1860

Qy   1861  AAAATGAACTCAAAAACCAAACGATTCGTTGGAGGAAATGGAGCATTTGATATGCCTGCA 1920
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1861  AAAATGAACTCAAAAACCAAACGATTCGTTGGAGGAAATGGAGCATTTGATATGCCTGCA 1920

Qy   1921  CTGGGATTGTAG 1932
           ||||||||||||
Db   1921  CTGGGATTGTAG 1932
```

```
RESULT 2
US-10-087-192-1663/c
; Sequence 1663, Application US/10087192
; Publication No. US20020182586A1
; GENERAL INFORMATION:
;     APPLICANT: Morris, David W.
;     APPLICANT: Engelhard, Eric K.
;     TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR
;     TITLE OF INVENTION: CANCER
;     FILE REFERENCE: 52945200012
;     CURRENT APPLICATION NUMBER: US/10/087,192
;     CURRENT FILING DATE: 2002-03-01
;     PRIOR APPLICATION NUMBER: US 09/747,377
;     PRIOR FILING DATE: 2000-12-22
;     PRIOR APPLICATION NUMBER: US 09/798,586
;     PRIOR FILING DATE: 2001-03-02
;     NUMBER OF SEQ ID NOS: 2059
;     SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1663
```

LENGTH: 193853
; TYPE: DNA
; ORGANISM: Mus musculus
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(193853)
; OTHER INFORMATION: n = A,T,C or G
US-10-087-192-1663

Query Match   2.6%;   Score 50.4;   DB 13;   Length 193853;
Best Local Similarity   28.9%;   Pred. No. 0.94;
Matches 123;   Conservative   0;   Mismatches 302;   Indels   0;   Gaps   0;

Qy   673 GGAAGAGGTGAAGAACAACGATGATGAGTGGACCAAAGAATGCAACAAAGA 732
              ||||| ||||||||| ||||||||||||||||||||||||||||||||
Db 34738 GAAAAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAA 34679

Qy   733 GAACTTCAAGAGAGAAGAGATGAAGATGATGATGAAGATCTTGAGGACGAGGA 792
              ||| |||||| ||||||
Db 34678 GAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAAAGAGAGA 34619

Qy   793 AGAAGTTCGGATGGAGAACCACAAAGTGAAGCAGAACCACAAGAGAGATTAGCCAGG 852
              |||||
Db 34618 AGAAGNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 34559

Qy   853 AGATTGAAAAGTAGTCCTAGATTAAAAGAGCTTTTACAGAATGCGGAGTTCAATCATTG 912
Db 34558 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 34499

Qy   913 CTCTCTTACCAACGAATGAGGGATTCTCCACTGAGCAAGGCCTTGGCTATGAAC 972
Db 34498 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 34439

Qy   973 GATGAGGATGAAAGTGCATTCCGCGCAATGCGAGGCTCGTGCAAAACTAGATCAAAAATCT 1032
              ||| ||||||
Db 34438 NNNNNNAAGGAGGAGGAGGAGGAGGAGGAAGAGGAAGAAGAGAGAGAGGAA 34379

Qy   1033 CAACTTGTGCTCCGGTCTCCATGGTTTTCGGAGAGTCTGATGATGAAGACGAAGAAGAT 1092
              ||||||| || || |||
Db 34378 GAGGAAGAGGAAGAGGAAGAAGAGGAAGAGGAAGAAGAGGAAGAAGAGGAA 34319

Qy   1093 GAAAA 1097
              |||
Db 34318 GAAGA 34314

Query Match   2.4%;   Score 47.2;   DB 17;   Length 3534;
Best Local Similarity   46.0%;   Pred. No. 0.64;
Matches 160;   Conservative   0;   Mismatches 188;   Indels   0;   Gaps   0;

Qy   1174 TCAAATGATGAATTGAAAAGTGCATTGGACAGAATTAAATATCGAGTTGATGACGTGGAA 1233
              ||||||||||||||| | ||||||||||| |||||||||||| ||||| |||
Db   1204 TCTAAGGAGTAAAATTAAAGAAGAAGAAATAAATTCATTAAAAATAGATGGAGACAGA 1263

Qy   1234 AAGTATCTTGCTCCAAAGCCGATGGAATTCAATCCAAAACCTCAGCCTCGGCTACTTTGCT 1293
              ||| ||| | ||||| | ||||| |||||| | |||| || |||||||| ||
Db   1264 AGAGAAAAGATACAAAAACTATATGAAAATAGATGAAATTAAAGATTAAAATTTAGAA 1323

Qy   1294 CCACGTAAAAATCCCAACAGACCACGTAAAATGCTTCCATTATTAATTGATCTGATCCA 1353
              || |||||| ||| || ||| || |||||||| ||||| |||| ||| ||||
Db   1324 GTTTGCAATATAACAGAAAGAATAAATATAAAATAAAAGCTTAAAACGATGATGAATTA 1383

Qy   1354 AAAGTTCAAGAGACGAAATACGAAGACATCCAAGTACCAGTAAAATTGCAAAAGAATCA 1413
              ||||||| |||||| ||| |||||| ||||| || ||| ||| |||
Db   1384 AAGTATAAAGATACATTAGAACAACAAAGAGATATAAAAAATTAGAATATGTTTCA 1443

Qy   1414 AGAGTTTTGACAAATTTGAAGAATAATCCAAGTCTTGCTGCATTGTTCATGGATGATAAA 1473
              |||||| |||||| || |||||| |||| || ||| ||||| ||| |||| ||||||
Db   1444 GGCAAAAGAGAAGAGATTTAATAAAAAACTCTCCTCGTAACAATGATATTTTGCTTGATAAA 1503

Qy   1474 TTAGAGATACATTGAAAAGGAAGGCAAAATGTTAACTGATGAACAGAAA 1521
              ||||| |||||| |||| | |||||| |||| |||||| ||||||
Db   1504 AAAGATTATATAAATAAATAATTAATGGAAATTTTAATGATTTAAATAAA 1551

```
;     LOCATION: (1)...(96602)
;     OTHER INFORMATION: n = any nucleotide
US-10-085-117-61

Query Match             2.4%;  Score 46.2;  DB 17;  Length 96602;
Best Local Similarity   52.9%;  Pred. No. 7.6;
Matches    99;  Conservative    0;  Mismatches   88;  Indels    0;  Gaps    0;

Qy        635 AAAGAGCAAATGTTAGGTAAAAATGCCCCAGTTGCCGGTGGAAGAGTGAAGAACAACGGA    694
              |||||| |||  ||| | |   ||||||   |||| || ||||||||||||| ||||||||
Db      50809 AAGGAGAAGAAGAAGAAGAAGAAGAAGAAGGAGAAGGAGAAGGAGAAGAAGAAGAAGAAG  50750

Qy        695 TGATGATGAATCGAGTGGACCAAAGAATGCAACAAAGAGAACTTCAAGAGAGAAGATGAAG    754
              |||   |     ||  ||   |||   ||| |||  ||||| || |||| |||||| ||||
Db      50749 AAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAG  50690

Qy        755 ATGATGATGATCTTGAGGACGAGGATGTACCCAGAGAAGAAGTTCGGATGGAGAACCAC    814
              |   |  ||| |   || || ||| ||  |||| ||||||||||||||||||| ||||||
Db      50689 AAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAAGAGAAAAGAGAGCAGCAG  50630

Qy        815 AAAGTGA    821
              ||||| |
Db      50629 AAGCAGA  50623


RESULT 5
US-10-239-676-28/c
;  Sequence 28, Application US/10239676
;  Publication No. US20030082609A1
; GENERAL INFORMATION:
;   APPLICANT: OLEK, Alexander
;   APPLICANT: PIEPENBROCK, Christian
;   APPLICANT: BERLIN, Kurt
;   TITLE OF INVENTION: Diagnosis of Diseases Associated with Gene Regulation
;   FILE REFERENCE: 5013.1003
;   CURRENT APPLICATION NUMBER: US/10/239,676
;   CURRENT FILING DATE: 2002-09-24
;   PRIOR APPLICATION NUMBER: PCT/EP01/03968
;   PRIOR FILING DATE: 2001-04-06
;   DE 10019058.8
;                2000-04-06
;   DE 10019173.8
;                2000-04-07
;   DE 10032529.7
;                2000-06-30
;   DE 10043826.1
;                2000-09-01
; NUMBER OF SEQ ID NOS: 228
; SEQ ID NO 28
;   LENGTH: 17848
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (6075, 8510, 8516, 8680, 9019, 15666)
US-10-239-676-28

Query Match             2.4%;  Score 45.8;  DB 14;  Length 17848;
Best Local Similarity   48.6%;  Pred. No. 3.7;
Matches   125;  Conservative    0;  Mismatches  132;  Indels    0;  Gaps    0;

Qy       1340 TTGGATCTGATCCAAAAGTTCAAGAGGAAATACGAAGACATCCAAGTACCGAATGAAAA   1399
              ||||||||||||||||||||||||||||| ||  ||||||||||||||| ||  ||||||
Db       7119 TTATAACTATTAAAATAATTAAAAAAAAAATACAACCTAAATAAAAAAAACTAAAA   7060

Qy       1400 TTGCAAAAGAATCAAGAGTTTTGACAAATTTGAAGAATAATCCAAGTCTTGCTGCATTGT   1459
              |||||| |  |||||||  ||| |||||| ||||||  |  ||   |  |||| |||||
Db       7059 TAACAAAAAAATAACGATTTCAAAAAATTTTAAAAACGAAATTAACAAAATATAATAAC   7000

Qy       1460 TCATGGATGATAAATTAGAGAATACATTGAAAGGAAGGCAAATGTTAACTGATGAACAGA   1519
              ||||| |||||  |   ||| || ||||  |  |||   |||  |||| |||| ||||||
Db       6999 CACTTAAAAATAAAATAAAAAAAAAATAACACTAAAATTCATAAAAAATA   6940
```

```
Db       6999 CACTTAAAAATAAAATAAAAAAAAAATAACACTAAAATTCATAAAAAATA   6940

Qy       1520 AAGGTAGAACACGTGTCAAAACAATTCGTGGTGCCACCACTG   1579
              ||||| ||||||||||||||||||||||||| ||  ||||||
Db       6939 AAATAAAAAAAAAACGATCTAAAAAAATTAAAATAATATATTTTATATAAAAATACCTA   6880

Qy       1580 CAAAAGCTGAAATGATT   1596
              |||| ||||||||
Db       6879 TAAAACATAAAAATATT   6863


RESULT 6
US-10-240-453-38/c
;  Sequence 38, Application US/10240453
;  Publication No. US20030148326A1
; GENERAL INFORMATION:
;   APPLICANT: OLEK, Alexander
;   APPLICANT: PIEPENBROCK, Christian
;   APPLICANT: BERLIN, Kurt
;   TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA
;   TITLE OF INVENTION: Transcription
;   TITLE OF INVENTION: by Means of Assessing the Methylation Status of Genes Associated
;   TITLE OF INVENTION: With DNA TRanscription
;   FILE REFERENCE: 5013.1009
;   CURRENT APPLICATION NUMBER: US/10/240,453
;   CURRENT FILING DATE: 2002-10-02
;   PRIOR APPLICATION NUMBER: PCT/EP01/03973
;   PRIOR FILING DATE: 2001-04-06
;   PRIOR APPLICATION NUMBER: DE 10019058.8
;   PRIOR FILING DATE: 2000-04-06
;   PRIOR APPLICATION NUMBER: DE 10019173.8
;   PRIOR FILING DATE: 2000-04-07
;   PRIOR APPLICATION NUMBER: DE 10032529.7
;   PRIOR FILING DATE: 2000-06-30
;   PRIOR APPLICATION NUMBER: DE 10043826.1
;   PRIOR FILING DATE: 2000-09-01
;   NUMBER OF SEQ ID NOS: 350
; SEQ ID NO 38
;   LENGTH: 17848
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (6075, 8510, 8516, 8680, 9019, 15666)
US-10-240-453-38

Query Match             2.4%;  Score 45.8;  DB 15;  Length 17848;
Best Local Similarity   48.6%;  Pred. No. 3.7;
Matches   125;  Conservative    0;  Mismatches  132;  Indels    0;  Gaps    0;

Qy       1340 TTGGATCTGATCCAAAAGTTCAAGAGGAAATACGAAGACATCCAAGTACCGAATGAAAA   1399
              ||||||||||||||||||||||||||||| ||  ||||||||||||||| ||  ||||||
Db       7119 TTATAACTATTAAAATAATTAAAAAAAAAATACAACCTAAATAAAAAAAACTAAAA   7060

Qy       1400 TTGCAAAAGAATCAAGAGTTTTGACAAATTTGAAGAATAATCCAAGTCTTGCTGCATTGT   1459
              |||||| |  |||||||  ||| |||||| ||||||  |  ||   |  |||| |||||
Db       7059 TAACAAAAAAATAACGATTTCAAAAAATTTTAAAAACGAAATTAACAAAATATAATAAC   7000

Qy       1460 TCATGGATGATAAATTAGAGAATACATTGAAAGGAAGGCAAATGTTAACTGATGAACAGA   1519
              ||||| |||||  |   ||| || ||||  |  |||   |||  |||| |||| ||||||
Db       6999 CACTTAAAAATAAAATAAAAAAAAAATAACACTAAAATTCATAAAAAATA   6940

Qy       1520 AAGGTAGAACACGTGTCAAAACAATTCGTGGTGCCACCACTG   1579
              ||||| ||||||||||||||||||||||||| ||  ||||||
Db       6939 AAATAAAAAAAAAACGATCTAAAAAAATTAAAATAATATATTTTATATAAAAATACCTA   6880

Qy       1580 CAAAAGCTGAAATGATT   1596
              |||| ||||||||
Db       6879 TAAAACATAAAAATATT   6863
```

RESULT 7
US-10-257-166-58/c
; Sequence 58, Application US/10257166
; Publication No. US20040023230A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Method and Nucleic Acids for Analysing the Methylation of
; TITLE OF INVENTION: Genes Implicated in Pharmacogenomics
; FILE REFERENCE: 5013.1011
; CURRENT APPLICATION NUMBER: US/10/257,166
; CURRENT FILING DATE: 2002-10-07
; PRIOR APPLICATION NUMBER: PCT/EP01/07470
; DE 1003529.7
; DE 1043826.1
; PRIOR FILING DATE: 2001-06-29
; 2000-06-30
; 2000-09-01
; NUMBER OF SEQ ID NOS: 178
; SEQ ID NO 58
; LENGTH: 17848
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (6075, 8510, 8516, 8680, 9019, 15666)
US-10-257-166-58

Query Match   2.4%; Score 45.8; DB 17; Length 17848;
Best Local Similarity 48.6%; Pred. No. 3.7;
Matches 125; Conservative   0; Mismatches 132; Indels   0; Gaps   0;

Qy   1340 TTGGATCTGATCCAAAAGTTCAAGAGGAAATACGAAGACATCCAAGTACCGAATGGAAAA 1399
          |||||| || |||||| |||||| ||||| |||| |||| || |||| |||| || ||
Db   7119 TTATAACTATTAAAATAATTAAAAAAAATACAAACCTAAATAAAAAAAAACTAAAA 7060

Qy   1400 TTGCAAAAGAATCAAGAGTTTGACAAATTTGAAGAATAATCCAAGTCTTGCTGCATTGT 1459
          | ||||| ||| |||||||| |||||||| ||||| ||||| |||| || ||||
Db   7059 TAACAAAAAAATAACGAATTTCAAAAAATTTTAAAAACGAAATTAACAAAATATAATAAC 7000

Qy   1460 TCATGGATGATAAATTAGAGAATACATTGAAAGGAAGGCAAATGTTAACTGATGAACAGA 1519
          |||| ||| |||||| |||||| ||||| ||| ||||| |||| ||| || ||||
Db   6999 CACTTAAAATAAAATAAAAATAAAAAAAAAATAAACTAAAATTCATAAAAATA 6940

Qy   1520 AAGGTAGAACACGTGTCAAAACAATTCGTGCATTCCAAGACTGTTCCGTGCACCACTG 1579
          ||||| ||| ||||| ||| || ||||| |||| || || ||||
Db   6939 AAAATAAAAAAAAAACGATCTAAAAAAATTAAAATATAATATTTATATAAAATACCTA 6880

Qy   1580 CAAAAGCTGAAATGATT 1596
          ||| ||| ||| |||
Db   6879 TAAAACATAAAAATATT 6863

RESULT 8
US-10-367-094-176
; Sequence 176, Application US/10367094
; Publication No. US20040170982A1
; GENERAL INFORMATION:
; APPLICANT: David W. Morris
; APPLICANT: Marc Malandro
; TITLE OF INVENTION: Novel Therapeutic Targets in Cancer
; FILE REFERENCE: 529452001500
; CURRENT APPLICATION NUMBER: US/10/367,094
; CURRENT FILING DATE: 2003-02-14
; NUMBER OF SEQ ID NOS: 203
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 176
; LENGTH: 127767
; TYPE: DNA
; ORGANISM: Mus musculus
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(127767)
; OTHER INFORMATION: n = A,T,C or G
US-10-367-094-176

Query Match   2.4%; Score 45.8; DB 19; Length 127767;
Best Local Similarity 53.4%; Pred. No. 11;
Matches 118; Conservative   0; Mismatches 102; Indels   1; Gaps   1;

Qy   673 GGAAGAGGTGAAGAACAACGGATGATGATGAATGCAGTGGACCAAAGAATGCAACAAAGA 732
          |||||| ||||||| || |||||||||| |||||| |||||||||||| ||||||||||
Db 100909 GAAGGAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAA 100968

Qy   733 GAACTTCAAGAGAAGAAGATGAAGATGAAGA 792
          ||| |||||| ||||||||||||| |||||| |||||
Db 100969 GAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGA-AGAAAA 101027

Qy   793 AGAAGTTCGGATGAGAACCACAAAGTGAAGCAGCATCAGAGGAAGAGATTTAGCCAGG 852
          ||||| |||||| |||||||| ||||||||||||||| ||||||||| ||||||
Db 101028 AGAAGAGGAGGAGGAGGAGGAGGTAGGAGGAGGAGGAAGAGGAGAAAAGTTTCAGGT 101087

Qy   853 AGATTGAAAAGTAGTCCTAGATTAAAAGAGCTTTTACAGAA 893
          |||| ||| || ||||||| ||||| ||||||
Db 101088 CTCTTCATAGTAAGATATGTTTGAAATAGAAAATTAGGAAA 101128

RESULT 9
US-10-425-115-78908
; Sequence 78908, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 78908
; LENGTH: 548
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(548)
; OTHER INFORMATION: unsure at all n locations
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_171990C.1
US-10-425-115-78908

Query Match   2.4%; Score 45.6; DB 20; Length 548;
Best Local Similarity 54.9%; Pred. No. 0.57;
Matches   90; Conservative   0; Mismatches 74; Indels   0; Gaps   0;

Qy   680 GTGAAGAACAACGGATGATGATGAATGCAGTGGACCAAAGAATGCAACAAAGAGAACTTC 739
          ||||||| || |||||||||| |||||| ||||||||||| |||||||| |||||| |||
Db   291 GGGAACATGCTCAAATCATGGCAAGGTTAGTTGGACTCGAAAGGGAAGAAAAGGAAGCTG 350

Qy   740 AAGAGGAAGAAGATGAAGATGATGATCTTGAGGACGAGGATGTACCCAGAAGAAGAAGTT 799
          |||||||||||||||||| ||| || |||||||| |||||||||||||||||
Db   351 GAAATACTTGTGAAGATGAATGATGATGGCCACGATGATGGCGACGATCAGAGGAAGGAT 410

Qy   800 CGGATGAGAACCACAAAGTGAAGCAGCATCAGAGGAAGAGAT 843
          |||| |||| ||| |||||||| |||||||| || ||
Db   411 AGGATGGTGAGGAAGATAAGGAACCCGAAATGCTTTAAGTGAT 454

RESULT 10
US-09-864-761-18684/c

```
; Sequence 18684, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
;   APPLICANT: Penn, Sharron G.
;   APPLICANT: Rank, David R.
;   APPLICANT: Hanzel, David K.
;   APPLICANT: Chen, Wensheng
;   TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
;   TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
;   FILE REFERENCE: Aeomica-X-1
;   CURRENT APPLICATION NUMBER: US/09/864,761
;   CURRENT FILING DATE: 2001-05-23
;   PRIOR APPLICATION NUMBER: US 60/180,312
;   PRIOR FILING DATE: 2000-02-04
;   PRIOR APPLICATION NUMBER: US 60/207,456
;   PRIOR FILING DATE: 2000-05-26
;   PRIOR APPLICATION NUMBER: US 09/632,366
;   PRIOR FILING DATE: 2000-08-03
;   PRIOR APPLICATION NUMBER: GB 24263.6
;   PRIOR FILING DATE: 2000-10-04
;   PRIOR APPLICATION NUMBER: US 60/236,359
;   PRIOR FILING DATE: 2000-09-27
;   PRIOR APPLICATION NUMBER: PCT/US01/00666
;   PRIOR FILING DATE: 2001-01-30
;   PRIOR APPLICATION NUMBER: PCT/US01/00667
;   PRIOR FILING DATE: 2001-01-30
;   PRIOR APPLICATION NUMBER: PCT/US01/00664
;   PRIOR FILING DATE: 2001-01-30
;   PRIOR APPLICATION NUMBER: PCT/US01/00669
;   PRIOR FILING DATE: 2001-01-30
;   PRIOR APPLICATION NUMBER: PCT/US01/00665
;   PRIOR FILING DATE: 2001-01-30
;   PRIOR APPLICATION NUMBER: PCT/US01/00668
;   PRIOR FILING DATE: 2001-01-30
;   PRIOR APPLICATION NUMBER: PCT/US01/00663
;   PRIOR FILING DATE: 2001-01-30
;   PRIOR APPLICATION NUMBER: PCT/US01/00662
;   PRIOR FILING DATE: 2001-01-30
;   PRIOR APPLICATION NUMBER: PCT/US01/00661
;   PRIOR FILING DATE: 2001-01-30
;   PRIOR APPLICATION NUMBER: PCT/US01/00670
;   PRIOR FILING DATE: 2001-01-30
;   PRIOR APPLICATION NUMBER: US 60/234,687
;   PRIOR FILING DATE: 2000-09-21
;   PRIOR APPLICATION NUMBER: US 09/608,408
;   PRIOR FILING DATE: 2000-06-30
;   PRIOR APPLICATION NUMBER: US 09/774,203
;   PRIOR FILING DATE: 2001-01-29
;   NUMBER OF SEQ ID NOS: 49117
;   SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 18684
;   LENGTH: 272
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   OTHER INFORMATION: MAP TO AL034548.24
;   OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 3
;   OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.4
;   OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 2.1
;   OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 3.2
;   OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 2.1
;   OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 2.7
;   OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 2.1
;   OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 2.4
;   OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1.3
;   OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 1.7
US-09-864-761-18684

Query Match       2.3%;   Score 45.4;   DB 9;   Length 272;
Best Local Similarity 53.7%;   Pred. No. 0.43;
Matches  94;  Conservative  0;  Mismatches  81;  Indels  0;  Gaps  0;

Qy        707 GAGTGGACCAAAGAATGCAACAAAGAACTTCAAGGAGGAAGATGAAGATGATGATGATC 766
```

```
Db        222 GAGGAGGAGGAGGAGGAAGAGGAAGAGGAAGAGGAAGAGGAAGAGGAAGAGGAAGAAGA 163
               |||     |||    ||    |||  ||||||||||||   |||||||||||||||| |||

Qy        767 TTGAGGACGAGGATGTACCCAGAAGAAGAAGTTCGGATGAGGAAGATGAAGATGAAGCAG 826
               ||||   ||||||||    |||||||||||||||   |||||||||||||||||||||||

Db        162 AAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAG 103
               |||    ||||||||||||||||||||||||||||||||||||||||||||||||||||

Qy        827 AGCATCAGAGAAGAGAGATTTAGCCAGGAGATTGAAAAGTAGTCCTAGATTAAAAGA 881
               ||   |||||||   ||    |||||||||||||||||||||  ||||||

Db        102 AAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAGGAAGAAGAAGAAGAAAAGA 48
```

```
RESULT 11
US-10-723-860-199/c
; Sequence 199, Application US/10723860
; Publication No. US20040253606A1
; GENERAL INFORMATION:
;   APPLICANT: Aziz, Natasha
;   APPLICANT: Ginsburg, Wendy M.
;   APPLICANT: Zlotnik, Albert
;   TITLE OF INVENTION: Methods of Diagnosis of Soft Tissue Sarcoma, Compositions &
;   TITLE OF INVENTION: Methods for Screening for Soft Tissue Sarcoma Modulators
;   FILE REFERENCE: 05882.0193.NPUS01
;   CURRENT APPLICATION NUMBER: US/10/723,860
;   CURRENT FILING DATE:  2003-11-26
;   PRIOR APPLICATION NUMBER: 60/429,739
;   PRIOR FILING DATE: 2002-11-26
;   NUMBER OF SEQ ID NOS: 8393
;   SOFTWARE: PatentIn version 3.2
; SEQ ID NO 199
;   LENGTH: 153170
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-723-860-199

Query Match       2.3%;   Score 45.4;   DB 20;   Length 153170;
Best Local Similarity 53.7%;   Pred. No. 16;
Matches  94;  Conservative  0;  Mismatches  81;  Indels  0;  Gaps  0;

Qy        707 GAGTGGACCAAAGAATGCAACAAAGAACTTCAAGGAGGAAGATGAAGATGATGATGATC 766
               ||

Db      91029 GAGGAGGAGGAGGAGGAAGAGGAAGAGGAAGAGGAAGAGGAAGAGGAAGAGGAAGAAG 90970

Qy        767 TTGAGGACGAGGATGTACCCAGAAGAAGAAGTTCGGATGAGGAAGATGAAGATGAAGCAG 826
               |||                    |||||||  ||    ||||||

Db      90969 AAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAG 90910

Qy        827 AGCATCAGAGAAGAGAGATTTAGCCAGGAGATTGAAAAGTAGTCCTAGATTAAAAGA 881
               ||   ||||||||   ||    ||||||

Db      90909 AAGAAGAAGAAGAAGAAGAAGAAGAAGAGGAAGAAGAAGAAGAAGAAAAGA 90855
```

```
RESULT 12
US-10-085-117-334/c
; Sequence 334, Application US/10085117
; Publication No. US20030232334A1
; GENERAL INFORMATION:
;   APPLICANT: Morris, David W.
;   APPLICANT: Engelhard, Eric K.
;   TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR CANCER
;   FILE REFERENCE: 52945200.0121
;   CURRENT APPLICATION NUMBER: US/10/085,117
;   CURRENT FILING DATE:  2002-02-27
;   PRIOR APPLICATION NUMBER: US 09/798,586
;   PRIOR FILING DATE: 2001-03-02
;   NUMBER OF SEQ ID NOS: 361
;   SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 334
;   LENGTH: 154817
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: variation
```

```
;    LOCATION: (1)...(154817)
;    OTHER INFORMATION: n = any nucleotide
US-10-085-117-334

Query Match             2.3%;  Score 45;  DB 17;  Length 154817;
Best Local Similarity  50.2%;  Pred. No. 20;
Matches  111;  Conservative   0;  Mismatches  110;  Indels   0;  Gaps   0;

Qy     635 AAAGAGCAATGTTAGGTAAAAATGCCCAGTTGCCGGTGGAAGAGGTGAAGAGAACAACGGA 694
               || ||  |||||  |||||||  ||||||||||||||||||||||||||
Db   49908 AAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAGAAGAAGAAG 49849

Qy     695 TGATGATGAATCGAGTGACCAACAAAGAAGAACTTCAACGAGGAGGATGAAG 754
               ||  |  ||  |||||||||||  |||||  |||||||| ||||
Db   49848 AAGAAGAGGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAGAGGAGGAGG 49789

Qy     755 ATGATGATGATCTTGAGGACGAGGATGTACCCAGAAGAAGAAGAAGTTCGGATGGAGAACCAC 814
               ||  |||||    |  |||||||    ||||||||| |||||||    ||||
Db   49788 AGGAGGAGGAGGAGGAGGAGGAGGAGGAAAGAGGGAAAGAGGAGGAGGAAGAGGAGGAGGAG 49729

Qy     815 AAAGTGAAGCAGAGCATCAGAGAAGAGATTTAGCCAGGAGA 855
               || ||||||    | |  |||||||||  | ||||
Db   49728 AAGAGGAAGAGGAGGAGGAAGAAGAAGGAGGAGGAAGAAGGACGA 49688

RESULT 13
US-10-029-386-3612
;  Sequence 3612, Application US/10029386
;  Publication No. US2003019704A1
; GENERAL INFORMATION:
;    APPLICANT: Penn, Sharron G.
;    APPLICANT: Rank, David R.
;    APPLICANT: Hanzel, David K.
;    TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR G
;    TITLE OF INVENTION: EXPRESSION ANALYSIS TWO
;    FILE REFERENCE: AEOMICA-X-2
;    CURRENT APPLICATION NUMBER: US/10/029,386
;    CURRENT FILING DATE: 2001-12-20
;    NUMBER OF SEQ ID NOS: 34288
;    SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
;    SEQ ID NO 3612
;    LENGTH: 543
;    TYPE: DNA
;    ORGANISM: Homo sapiens
;    FEATURE:
;    OTHER INFORMATION: MAP TO CHR22_144.0
;    OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.31
;    OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1
;    OTHER INFORMATION: EST_HUMAN HIT: AI446528.1, EVALUE 7.60e-01
US-10-029-386-3612

Query Match             2.3%;  Score 44.8;  DB 16;  Length 543;
Best Local Similarity  50.5%;  Pred. No. 0.91;
Matches  109;  Conservative   0;  Mismatches  107;  Indels   0;  Gaps   0;

Qy     625 GCTCTTGCTCAAAGAGCAATGTTAGGTAAAAATGCCCCAGTTGCCGGTGGAAGAGGTGAA 684
               ||||| || ||||||||||||||||
Db     160 GCTCCCTTTAGAAGAGGAGGAGGAGGAGGAGGAGGAAGAGGAGGCCGGAGGAGGAA 219

Qy     685 GAACAACGGATGATGATGATCGAGTGACCAACAAAGAATGCGAACAAAGAACTTCAAGAG 744
               ||| ||||    |||||||||||
Db     220 GAAGAAGAAAGATGAGGAGGAAGAAGAGGAAGAAGATGAGGAGGAGGAGGAAGAAGA 279

Qy     745 GAAGATGAAGATGATGATGATCTTGAGGACGAGGATCTTGAGGACGACGAGTACCCAGAAGAAGAAGTTCGGAT 804
               ||||||||||| |||| |||| |||||
Db     280 GAAGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGAAGAAGAAGGAGGAGGAGGAGGAGGA 339

Qy     805 GGAGAACCACAAGTGAAGCAGAGCATCAGAGAAGA 840
               |||||  || || |||||||    | |
Db     340 AGAGGAGGAAGAAGAGGAAGGAGGAGGAGGAGAGA 375

RESULT 14
```

```
US-10-424-599-83149
;  Sequence 83149, Application US/10424599
;  Publication No. US2004003172A1
; GENERAL INFORMATION:
;    APPLICANT: La Rosa Thomas J
;    APPLICANT: Kovalic David K
;    APPLICANT: Zhou Yihua
;    APPLICANT: Cao Yongwei
;    TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
;    TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
;    FILE REFERENCE: 38-21(53223)B
;    CURRENT APPLICATION NUMBER: US/10/424,599
;    CURRENT FILING DATE: 2003-04-28
;    NUMBER OF SEQ ID NOS: 285684
;    SEQ ID NO 83149
;    LENGTH: 694
;    TYPE: DNA
;    ORGANISM: Glycine max
;    FEATURE:
;    OTHER INFORMATION: Clone ID: PAT_MRT3847_460C.1
US-10-424-599-83149

Query Match             2.3%;  Score 44.6;  DB 18;  Length 694;
Best Local Similarity  55.5%;  Pred. No. 1.2;
Matches  86;  Conservative   0;  Mismatches  69;  Indels   0;  Gaps   0;

Qy     168 GGGTGTTCAATTTGTTGATGCTCCATCAAAAAGGTCAAATGGAAATGCAAAAGGAGC 227
               |||||||||||||| || |||||||    ||||| ||||||||
Db     470 GGGTTTGCCATTAGCTCCTACAGTCCTTGGTTGCCATTTGTTTGATATGAAGATAATAGA 529

Qy     228 ATTTAAGACTCAATTAGAAGTTCTAGAGAAAGTACATCCTGATCAATTCGATCAAGTACAA 287
               ||| || |||| |||||||||||||||    |||||||||
Db     530 ATGGAAGACTGTATTAGACAAACTTAAGAGAATTCCTCATGATGATCAAGTACAAAAGAAGTT 589

Qy     288 AAAGCTAAAAGTTGATGATTTGGCAGCTGATGCAG 322
               ||| ||||| ||||||| |||
Db     590 AAAAATAAGCTATGATGGTTTAAGTGATGATACAG 624

RESULT 15
US-10-425-114-11335
;  Sequence 11335, Application US/10425114
;  Publication No. US2004003488A1
; GENERAL INFORMATION:
;    APPLICANT: Liu, Jingdong
;    APPLICANT: Zhou, Yihua
;    APPLICANT: Kovalic, David K.
;    APPLICANT: Screen, Steven E
;    APPLICANT: Tabaska, Jack E
;    APPLICANT: Cao, Yongwei
;    TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
;    TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
;    FILE REFERENCE: 38-21(53313)B
;    CURRENT APPLICATION NUMBER: US/10/425,114
;    CURRENT FILING DATE: 2003-04-28
;    NUMBER OF SEQ ID NOS: 73128
;    SEQ ID NO 11335
;    LENGTH: 694
;    TYPE: DNA
;    ORGANISM: Glycine max
;    FEATURE:
;    OTHER INFORMATION: Clone ID: 701053515_FLI
US-10-425-114-11335

Query Match             2.3%;  Score 44.6;  DB 18;  Length 694;
Best Local Similarity  55.5%;  Pred. No. 1.2;
Matches  86;  Conservative   0;  Mismatches  69;  Indels   0;  Gaps   0;

Qy     168 GGGTGTTCAATTTGTTGATGCTCCATCAAAAAGGTCAAATGGAAATGCAAAAGGAGC 227
               |||||||||||||| || |||||||    ||||| ||||||||
Db     470 GGGTTTGCCATTAGCTCCTACAGTCCTTGGTTGCCATTTGTTTGATATGAAGATAATAGA 529

Qy     228 ATTTAAGACTCAATTAGAAGTTCTAGAGAAAGTACATCCTGATCAATTCGATCAAGTACAA 287
```

```
Db         ||  ||||||  ||||||    ||   |||| |  | || |||||||| | |||  |
       530 ATGGAAGACTGTATTAGACAAACTTAAGAGAATTCCTCATGATCAAGTACAAAAGAAGTT 589

Qy     288 AAAGCTAAAAGTTGATGATTTGGCAGCTGATGCAG 322
           ||| |||||||| | |||||| |||
Db     590 AAAAATAAGCTATGATGGTTTAAGTGATGATACAG 624
```

Search completed: July 15, 2005, 15:41:45
Job time : 1226 secs

```
                    GenCore version 5.1.6
            Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on:        July 15, 2005, 10:38:45 ; Search time 6503 Seconds
                                       (without alignments)
                          11308.653 Million cell updates/sec

Title:         US-10-736-868-1
Perfect score: 1932
Sequence:      1 atgattatttattttt.......tgcctgcactgggattgtag 1932
Scoring table: IDENTITY NUC
               Gapop 10.0 , Gapext 1.0

Searched:      3239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters:    68479088

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :     EST:*
               1: gb_est1:*
               2: gb_est2:*
               3: gb_htc:*
               4: gb_est3:*
               5: gb_est4:*
               6: gb_est5:*
               7: gb_est6:*
               8: gb_gss1:*
               9: gb_gss2:*

        Pred. No. is the number of results predicted by chance to have a
        score greater than or equal to the score of the result being printed,
        and is derived by analysis of the total score distribution.
```

SUMMARIES

Result No.		Score	% Match	Query Length	DB	ID	Description
c	1	273	14.1	300	1	AV178298	AV178298 AV178298
c	2	124.6	6.4	476	1	BI741875	BI741875 kt82b03.y
c	3	57.4	3.0	1101	9	CNS0039G	AL063921 Drosophil
c	4	55.8	2.9	997	9	CNS005TE	CD167558 Drosophil
c	5	54.4	2.8	478	6	CD167558	CD167558 MM1-0013G
c	6	54.4	2.8	534	6	CD185097	CD185097 MS1-0051T
c	7	53.8	2.8	1359	9	AG310736	AG310736 Mus muscu
c	8	53.8	2.8	622	5	BU418103	BU418103 603959413
c	9	53	2.7	488	8	AZ379487	AZ379487 1M0134N11
c	10	53	2.7	1196	9	CNS0207H	AL206342 Tetraodon
c	11	50.8	2.6	610	9	CR089470	CR089470 Reverse s
c	12	50.8	2.6	838	9	CR214149	CR214149 Reverse s
c	13	50.8	2.6	1101	9	CNS0037Q	AL064465 Drosophil
c	14	50.8	2.6	1133	9	CL111125	CL111125 ISB1-54M2
c	15	50.6	2.6	1201	9	CNS02O3O	AL206205 Tetraodon
c	16	50.2	2.6	440	8	AZ408774	AZ408774 1M0180D24
c	17	50.2	2.6	572	8	AZ959862	AZ959862 2M0227M11
c	18	50	2.6	787	9	AG585416	AG585416 Mus muscu
c	19	49.8	2.6	550	8	BH264512	BH264512 CH230-121
c	20	49.6	2.6	501	6	CD167504	CD167504 MM1-0013G
c	21	49.6	2.6	562	9	CE632047	CE632047 tigr-gss-
c	22	49.2	2.5	637	9	CE379668	CE379668 tigr-gss-
c	23	49	2.5	744	9	AG447497	AG447497 Mus muscu
c	24	48.8	2.5	835	8	B2206628	B2206628 CH230-390
c	25	48.6	2.5	356	8	BZ118849	BZ118849 CH230-422
c	26	48.6	2.5	867	9	BZ270082	BZ270082 CH230-327
c	27	48.4	2.5	1101	9	CNS012W2	AL102092 Drosophil
c	28	48.2	2.5	635	9	CE470957	CE470957 tigr-gss-
	29	48	2.5	391	6	BY650247	BY650247 BY650247
	30	48	2.5	626	9	BJ332286	BJ332286 BJ332286
	31	48	2.5	654	9	CE797922	CE797922 tigr-gss-
c	32	47.8	2.5	1424	9	AG346078	AG346078 Mus muscu
c	33	47.6	2.5	578	8	AZ263752	AZ263752 RPCI-23-1
c	34	47.6	2.5	745	8	BH034701	BH034701 RPCI-24-2
c	35	47.6	2.5	1378	9	AG290218	AG290218 Mus muscu
	36	47.4	2.5	482	9	CE325476	CE325476 tigr-gss-
	37	47.4	2.5	521	9	CE203111	CE203111 tigr-gss-
	38	47.4	2.5	571	9	CE769781	CE769781 tigr-gss-
c	39	47.4	2.5	1067	9	CNS00C08	AL058564 Drosophil
c	40	47.2	2.4	2129	9	AG393084	AG393084 Mus muscu
c	41	47	2.4	754	8	AG564467	AG564467 Mus muscu
c	42	47	2.4	228	1	AU260061	AU260061 AU260061
c	43	47	2.4	323	4	BI076776	BI076776 L0256C10-
c	44	47	2.4	346	4	BG087750	BG087750 H3144E02-
c	45	46.8	2.4	676	7	CF696863	CF696863 CCAAB79TF

ALIGNMENTS

```
RESULT 1
AV178298/c
LOCUS       AV178298              300 bp    mRNA    linear   EST 21-JUL-1999
DEFINITION  AV178298 Yuji Kohara unpublished cDNA:Strain N2 hermaphrodite
            embryo Caenorhabditis elegans cDNA clone yk563c9 3', mRNA sequence.
ACCESSION   AV178298
VERSION     AV178298.1  GI:5558199
KEYWORDS    EST.
SOURCE      Caenorhabditis elegans
  ORGANISM  Caenorhabditis elegans
            Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
            Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE   1  (bases 1 to 300)
  AUTHORS   Kohara,Y., Shin-i,T., Thierry-Mieg,J., Thierry-Mieg,D., Mitsuki,H.,
            Nishigaki,A., Motohashi,T., Zeng,Q., Watanabe,H., Sugimoto,A.,
            Sano,M., Miyata,A., Mitani,Y., Iida,K., Uesugi,H., Sugiyama,Y. and
            Nomoto,H.
  TITLE     Expressed genes in C.elegans
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Yuji Kohara
            Genome Biology Lab.
            National Institute of Genetics
            Yata 1111, Mishima, Shizuoka 411, Japan
            Tel: 81-559-81-6854
            Fax: 81-559-81-6855
            Email: ykohara@lab.nig.ac.jp.
FEATURES             Location/Qualifiers
     source          1..300
                     /organism="Caenorhabditis elegans"
                     /mol_type="mRNA"
                     /strain="N2"
                     /db_xref="taxon:6239"
                     /clone="yk563c9"
                     /sex="hermaphrodite"
                     /dev_stage="embryo"
                     /clone_lib="Yuji Kohara unpublished cDNA:Strain N2
                     hermaphrodite embryo"
ORIGIN

Query Match             14.1%;  Score 273;  DB 1;  Length 300;
Best Local Similarity   99.3%;  Pred. No. 4.4e-61;
Matches  273;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy   1658 AGGAAGGCATACGAGATTGAGATGGACTGGAGCAAATGAAAAAGAAATTCCAGGACTTG 1717
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    300 AGGAAGGCATACGAGATTGAGATGGACTGGAGCAAATGAAAAAGAAATTCCAGGACTTG 241
```

```
Qy   1718  GAAGTCGCTTCATTCTCCCATCTCTTGATCCAACTATGCCAGCCTGAAACACGGCTTCT 1777
           ||||||||||||||||||| ||||||||||||||||||||||| ||||| |||||||||||
Db   240   GAAGTCGCTTCATTCTCCCAATCTTGATCCAACTATGCCANCCTTGAACACGGCTTTCT 181

Qy   1778  CGACTCAGGGGCGAGCCCGTGACGAGTGGGATACCATGTTCAAAATCCGAATAACTGGA 1837
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   180   CGACTCAGGGGCGAGCCCGTGACGAGTGGGATACCATGTTCAAAATCCGAATAACTGGA 121

Qy   1838  ATCCTGAGATGAAGTGGGTTCAAAATGAACTCAAAAACCAAACGATTCGTTGGAGGAA 1897
           ||||||||||| |||||||||||||||||||||||||||||||||||||||||||||||
Db   120   ATCCTGAGATCAAGTGGGTTCAAAATGAACTCAAAAACCAAACGATTCGTTGGAGGAA 61

Qy   1898  ATGGAGGACATTGATATGCCTGCACTGGGATTGTAG 1932
           ||||||||||||||||||||||||||||||||||||
Db   60    ATGGAGGACATTGATATGCCTGCACTGGGATTGTAG 26
```

RESULT 2
BI741875
LOCUS   BI741875   476 bp   mRNA   linear   EST 25-SEP-2001
DEFINITION kt82b03.y1 Strongyloides ratti L1 pAMP1 v3 Chiapelli McCarter
            Strongyloides ratti cDNA 5' similar to TR:P91120 P91120 SIMILAR TO
            HMG-BOX SINGLE-STRANDED DNA AND RNA BINDING PROTEINS. [1] ;, mRNA
            sequence.
ACCESSION   BI741875
VERSION   BI741875.1 GI:15742831
KEYWORDS   EST.
SOURCE   Strongyloides ratti
  ORGANISM Strongyloides ratti
            Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
            Panagrolaimoidea; Strongyloididae; Strongyloides.
REFERENCE   1 (bases 1 to 476)
  AUTHORS   McCarter,J., Clifton,S., Chiapelli,B., Pape,D., Martin,J.,
            Wylie,T., Dante,M., Marra,M., Hillier,L., Kucaba,T., Theising,B.,
            Bowers,Y., Gibbons,M., Ritter,E., Bennett,J., Franklin,C.,
            Tsagareishvili,R., Ronko,I., Kennedy,S., Maguire,L., Beck,C.,
            Underwood,K., Steptoe,M., Allen,M., Person,B., Swaller,T.,
            Harvey,N., Schurk,R., Kohn,S., Shin,T., Jackson,Y., Cardenas,M.,
            McCann,R., Waterston,R. and Wilson,R.
  TITLE   The Washington Univ. Nematode EST Project, 1999
  JOURNAL   Unpublished (1999)
COMMENT   Contact: McCarter JP
            The Washington Univ. Nematode EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            The library was constructed by Brandi Chiapelli and Dr. James
            McCarter (bchiapel@watson.wustl.edu & jmccarte@watson.wustl.edu) at
            Washington University, St. Louis. DNA Sequencing by: Washington
            University Genome Sequencing Center St. Louis.
            Seq primer: -40RP from Gibco
            High quality sequence stop: 306.
FEATURES   Location/Qualifiers
     source   1..476
                     /organism="Strongyloides ratti"
                     /mol_type="mRNA"
                     /db_xref="taxon:34506"
                     /dev_stage="L1"
                     /lab_host="DH10B"
                     /clone_lib="Strongyloides ratti L1 pAMP1 v3 Chiapelli
                     McCarter"
                     /note="Vector: pAMP1 (Gibco); The library was constructed
                     by Brandi Chiapelli and Dr. James McCarter at Washington
                     University, St. Louis. The cDNA was made by using
                     Dynabead oligo-dT priming (Dynal). PCR based library
                     using a modified protocol from the SMART PCR cDNA
                     Synthesis Kit from Clontech. Directionally cloned into the
                     UDG sites of pAMP1. Nematodes were provided by Dr. Mark
                     Viney of Bristol, UK."
ORIGIN
```

```
Query Match              6.4%;   Score 124.6;  DB 4;   Length 476;
Best Local Similarity   63.5%;   Pred. No. 1.3e-21;
Matches  209;  Conservative   0; Mismatches 114;  Indels 6;  Gaps 1;

Qy   1523  GTAGAACACGTGTCAAAACAATTCGTGCATTACCAAGACTGTTCGGTGC------ACCAA 1576
           ||||||||  ||||    ||| |||   |||||||  |||| ||||
Db   85    GTGGTACACGTGGAAATGCAATTACACATATACCACGTATGTGCTAAATTATTG 144

Qy   1577  CTGCAAAAGCTGAAATGATTGATGCAAAGGTATTCCAAGATATTGAAGAACGTCCCATTC 1636
           ||| ||  ||  |||||  ||| |||||    ||   |||||||||| |||
Db   145   GTGATGAACCAAATATTTATGATAATAAAACAAAACAAATTGTTGAGGAAGAGAAATTC 204

Qy   1637  CTCCATTGTTCTTGAACCAAAAGAAGGAGGCATACGAGATTGAGATGGGACTGGACAAATG 1696
           ||||| ||  ||| | |||||||  ||||| ||||||| |||||| ||||| ||||  ||
Db   205   CTCCACTTTTTTTTGTACCAAAAGGAAAACATACAAGATTAAGGTGGGTAACAGCAACAG 264

Qy   1697  AAAAGAAATTCCAGGAACTTGGAAGTCGCTTCATTCTCCCATCTCTTGATCCAACTATGC 1756
           || ||||||||| |||| |||||||| |  ||  ||  || | ||||||||||||| |||
Db   265   AACAAGAAATTCCCGGAATTGGAAGTAGATTTATTATTATACCACACTTGATCCAACTAGGC 324

Qy   1757  CAGCCTGAACACGGCTTTCTCGACTCAGGGGCCGAGCCCGTGACGAGTGGGATACCATGT 1816
           ||||  ||| ||  |  | |  || ||   ||| |  |||| || |||| ||||| |||
Db   325   CAGCAATTAATTCAGTTGTATCAACACAAGGAAAGAAATGAATATGAAACAACTT 384

Qy   1817  TCAAAATCCGAATAACTGGAATCCTGA 1845
           ||||| ||| ||| ||| || ||||||
Db   385   GGAAAATACCAAATGCTTGGAAATCTGA 413
```

RESULT 3
CNS0039G/c
LOCUS   CNS0039G   1101 bp   DNA   linear   GSS 03-JUN-1999
DEFINITION Drosophila melanogaster genome survey sequence TET3 end of BAC #
            BACR08K10 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL063921
VERSION   AL063921.1 GI:4941778
KEYWORDS   GSS.
SOURCE   Drosophila melanogaster (fruit fly)
  ORGANISM Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1 (bases 1 to 1101)
  AUTHORS   Genoscope.
  TITLE   Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT   Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES   Location/Qualifiers
     source   1..1101
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
                     /clone="BACR08K10"
                     /clone_lib="RPCI-98"
                     /note="end : TET3"
ORIGIN
```

melanogaster genome using these BACs. For further information
please see http://www.fruitfly.org The BDGP Drosophila
melanogaster BAC library was prepared by Kazutoyo Osoegawa and
Aaron Mammoser in Pieter de Jong's laboratory in the Department of
Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
NY. The library is named RPCI-98 and was constructed by partial
EcoRI digestion of Drosophila DNA provided by the BDGP from the
isogenic strain y2; cn bw sp, the same strain used for the BDGP's
P1 and EST libraries. A more detailed description of the library
and how to order individual BAC clones, the entire library, or
filters for hybridization from the BACPAC Resource Center can be
found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

```
FEATURES             Location/Qualifiers
     source          1..997
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
                     /clone="BACR12K22"
                     /clone_lib="RPCI-98"
                     /note="end : TET3"

ORIGIN
```

Query Match　　　2.9%;　Score 55.8;　DB 9;　Length 997;
Best Local Similarity　18.6%;　Pred. No. 0.0029;
Matches　79;　Conservative 154;　Mismatches 192;　Indels　0;　Gaps　0;

```
Qy    673 GGAGGAGGTGAAGAACAACACGGATGATGATGAATGCAATGCGAGTGGACCAAAGATGCAACAAAGA 732
          ::|:::|:. ::::|:||:.|:::|:|||:: :|.:|:||:|:||.:|||.|:::||.:.:|:||:|
Db    996 RRARRRAGARARARAARGAGARRRRGRRGARGARARARARRRRARGARAGARAAARAGARRRAGA 937

Qy    733 GAACTTCAAGAGGAAGATGAAGATGATGATCTTGAGGACGAGGATGTACCCAGAAGA 792
          |: .|:|:|::|:|::||::|:|:||:|:||::|:|.|||:|:||:||.::|:||.:
Db    936 GARARAAARRRRARRRRAGRRAGRAGRGRSAGGRGRGGRRAAAARDARARARARARA 877

Qy    793 AGAAGTTCGGATGGAGAACCACAAAGTGAAGCAGAGCATCAGACGAAGAGATTTAGCCAGG 852
          :.:|:||:::|.||:||||.:|::|||:|:||:|::|:.:::|||::||||.::|||.:||
Db    876 RRRARAKAAARAGARRRRGRRRRGRGRRGRGAGRRRAGAGGRAGRGRGRRRARARRRRAGAA 817

Qy    853 AGATTGAAAAGTAGTCCTAGATTAAAAGAGCTTTTACAGAATGGGGAAGTTCAATCATTG 912
          :.::|:||||:.:|.::|||::|||||||:||::::::.||::||::|.:|:|:||:.||
Db    816 RARARGRRARAARARRRRAGRRRRRARGARGRRRGGRRGRRRGRRRRAARAR 757

Qy    913 CTCTCTTACCAACGAATGAGGGATTCTCCACTGACGAAGCAAGCGAAGGCCTTTGGCTATGAAC 972
          .::.|:|||::| |:||||||.::: .|: .:.|.:||:||||||:.::|:::||.|||.:||
Db    756 ARARAGARAAARRRRAGARAGAAAAARARARRRRRRRGAGARARRARRRARAAAARAGAGAGA 697

Qy    973 GATGAGGATGAAAGTGCATTCCCGCCAATGGAGGCTCCGTCGCAAAACTAGATCAAAAATCT 1032
          |:||||||::||.::|:|::||||::|:||||:|.:||.:|||:|:||.||||::||.|||:
Db    696 RRRGRGRGAGARRGRGRGRGRGRGGARRRRRRGRRMRTRRARRRRRARGAGARRRRAR 637

Qy   1033 CAACTTGTGCTCCGGTCTCCATGGTTTTGGAGAGTCTGATGATGATGAAGACGAAGAGAT 1092
          |..|:|:|:::|:|::|:::.|||.::::||||:|:||:||:||:||:|||:||||||:|
Db    636 GARRRRRGRGRRGRARGARRRRRRAGRRRRGAGARRRRRRRRRRRAGAGAR 577

Qy   1093 GAAAA 1097
          : :::
Db    576 RGRRR 572
```

RESULT 5
CD167558/c

Query Match　　　3.0%;　Score 57.4;　DB 9;　Length 1101;
Best Local Similarity　15.3%;　Pred. No. 0.0011;
Matches　92;　Conservative 262;　Mismatches 247;　Indels　0;　Gaps　0;

```
Qy    677 GAGGTGAAGAACAACGGATGATGATGAATGCAATGCGAGTGGACCAAAGATGCAACAAAGAAC 736
          :|::|:.::: :.:::::|||:::|:::|||::::: :|:::: :::::::::::.::::::
Db   1100 KARRWGDDTWDRDTRKODWDWTKWWTWWKDRADDRRWAGDADRRWAWDDGAGTWWTATWWW 1041

Qy    737 TTCAAGAGGAAGATGAAGATGATGATCTTGAGGACGACGATGTACCCAGAAGAGAA 796
          :: ::.|: .::|| ||::|: || :|::::|:: |: :||::: .:||:|: |||
Db   1040 WWWWATWDTWWDKWWWWATAAKTDTAWTWWRTAWRADWAGRDRGAGKRDRDAATDADGA 981

Qy    797 GTTCGGATGGAGAACCACAAAGTGAAGCAGAGCATCAGAGAAGAGATTTAGCCAGGAGAT 856
          |::|::|::|:||.::||.:|||:|:||.::|:.::||.||::||||.::|::|:|:||||
Db    980 GRRDGGRRKRDKKDKRKDGDDDKKGGKKKKCAAAKAWTKWWDDWDWDWDKWKWDGAKDRKA 921

Qy    857 TGAAAAGTAGTCCTAGATTAAAAGAGCTTTTACAGAATGCGGAAGTTCAATCATTGCTCT 916
          :::: ::|.:|: ::|.||.:.:|.|::::::.||||:|:|::||::|:|::||::|::|:
Db    920 DDDDGAGDKDDGKGKDADDDTDGTKDDDKDWDDWDKAKGTWGDATWAWAATDWWWWG 861

Qy    917 CCTTACCAACGAATGAGGGATTCTCCACTGACGCAGCGAAGGCCTTTGGCTATGAACGATG 976
          .:::|::|::.:|:|: ::.||:::|::|:|.:|::||:|||:::|||:::|:||.|:::|
Db    860 WADADWWTWDAAADDWWADDRWDAWAWKWDDAWAWGARTADRRDWGDRAGKRGGARKRRD 801

Qy    977 AGGATGAAAGTGCATTCCGCGCAATGGAGGCTCGTGCAAAACTAGATCAAAAATCTCAAC 1036
          ::|::::: ::|:|::||:::|:|::|||::|:|:::.::.|||.||:::||:||:|:|||
Db    800 RKRADDKRDAADDRRDDAATWTWTTTTTRDTDDWKWKTDTWTRWAADRTWDRDDDDDRDR 741

Qy   1037 TTGTGCTCCGGTCTCCATGGTTTTGGAGAGTCTGATGATGAAGACGAAGAGAGATGAAA 1096
          :::|:::|:::|:||::.|||::::||||||:||:||:||:|:|| :|:||:|:||:|||
Db    740 AGTAGRKWRRTWKRRWKRRDTRWDDADADDTARDDRRRRGDDGADAGKGKKTGRKRRRRD 681

Qy   1097 ATTTGATTGATCCATCTGAAAATTCATTCCGTCGTCCACCACTTCGTCTTCTTCCGGAT 1156
          :|::|:|::|:::.:|::| ||::|::||::||:||::|::|::::|::|::::|::.|:
Db    680 RATWDRTDAWTWADAAWTTTDTDTDDWDKRDRRRKGARRRRRTTARAAWDWWTWKAWDWA 621

Qy   1157 TCGTTGAGAAATTGAAAGTCAAATGATGAATTGAAAAGTGCATTGGACAGAATTAAAATC 1216
          ||::|:|||:|.:|||:|:||::|||:|:::::||::||:|:||:::||:||:::|||.||
Db    620 KWDWKTRADRWDRWAADTWTDARKADRDWAKARAWRARRDRARAARADRRWTTKGKTTTA 561

Qy   1217 GAGTTGATGACGTGGAAAAGTATCTTGCTCCAAAGCCGATGAATTCCAAAACCTC 1276
          ::|:|:|.|::|::::::::|::.|::||::::||||::|:||:::|::|::|::|
Db    560 TWTTWARAAWWWAWWAWWWATTTATWTTWTTWTWTWTTTTTWTTTTAAWWWAAWWTATW 501

Qy   1277 A 1277

Db    500 A 500
```

RESULT 4
CNS005TE/c

AUTHORS   Verjovski-Almeida,S., DeMarco,R., Martins,E.A.L., Guimaraes,P.E.M.,
           Ojopi,E.P.B., Paquola,A.C.M., Piazza,J.P., Nishiyama,M.Y. Jr.,
           Kitajima,J.P., Adamson,R.E., Ashton,P.D., Bonaldo,M.F.,
           Coulson,P.S., Dillon,G.P., Farias,L.P., Gregorio,S.P., Ho,P.L.,
           Leite,R.A., Malaquias,L.C.C., Marques,R.C.P., Miyasato,P.A.,
           Nascimento,A.L.T.O., Ohlweiler,F.P., Reis,E.M., Ribeiro,M.A.,
           Sa,R.G., Stukart,G.C., Soares,M.B., Gargioni,C., Kawano,T.,
           Rodrigues,V., Madeira,A.M.B.N., Wilson,R.A., Menck,C.F.M.,
           Setubal,J.C., Leite,L.C.C. and Dias-Neto,E.
TITLE   Transcriptome analysis of the acoelomate human parasite Schistosoma
           mansoni
JOURNAL   Nat. Genet. 35 (2), 148-157 (2003)
MEDLINE   22879926
PUBMED   12973350
COMMENT   Contact: Dr. Sergio Verjovski-Almeida
           Departamento de Bioquimica
           Instituto de Quimica - Universidade de Sao Paulo
           Av. Prof. Lineu Prestes 748 sala 1200, 05508-900 Sao Paulo - SP,
           Brasil
           Tel: +55-11-3091-2173
           Fax: +55-11-3091-2186
           Email: verjo@iq.usp.br
           This sequence was derived from the FAPESP Schistosoma mansoni EST
           Genome Project. All sequences in the project were assembled and
           annotated. This entry and all the assembled sequences can be seen
           in the following URL http://bioinfo.iq.usp.br/schisto/
           Plate: MS1-0051T-L290 row: 6 column: A.
FEATURES   Location/Qualifiers
     source   1..534
                     /organism="Schistosoma mansoni"
                     /mol_type="mRNA"
                     /db_xref="taxon:6183"
                     /clone="MS1-0051T-L290-A06.B"
                     /sex="mixed pool"
                     /dev_stage="schistosomulum"
                     /lab_host="in vitro culture"
                     /clone_lib="MS1-0051"
                     /note="Vector: pGEM T-easy"
ORIGIN

Query Match   2.8%;   Score 54.4; DB 6;   Length 534;
Best Local Similarity 52.2%;   Pred. No. 0.0058;
Matches 121; Conservative   0; Mismatches 111; Indels   0; Gaps   0;

Qy   616 TTCCAACAAGCTCTTGCTCAAAGAGACAATGTTAGTAAAAATGCCCCAGTTGCCGGTGGA 675
          || |||| |||| ||||| ||||| || ||||| |||| |||||| |||||||
Db   183 TTAAAACACTTCCGTACTTCAACAGCGAGCTCTGAGAAAACTTCGAGCGAACCTACTGCA 242

Qy   676 AGAGGTGAAGAACAACGGATGATGAATCGAGTGGACCAAAGAATGCAACAAAGAGAA 735
          |||| ||||||| ||| ||| || |||||||||| | | |||| ||||||||
Db   243 CCAGCTGATGAACAACAGACAACAGACACGCCCGAAACAAACTGAACAAAGTGTTGACGAAACCTCA 302

Qy   736 CTTCAAGAGGAAGATGAAGATGATGAATCGTTGAGGACGAGGATGTACCCAGAAGAAGA 795
          | |||| ||||| ||| ||| |||| ||| |||||| || |||| ||||||||||
Db   303 CCTAAAGATGAAACTAAAGAAAAGATCAGCCTGTTGAACCAAATGAAAAAAAGAAATC 362

Qy   796 AGTTCGGATGGAGAACCACAAAGTGAAGCAGCAGAGCATCAGAGAAGAAGAGATTTAG 847
          ||||| |||||||||||||||||| | |||||||||| || || |||||||| |||
Db   363 CAAGAAATGTTGAAGTTGTAGGAGTAACAGAGCAACCACAACCATCTGTAG 414

RESULT 7
AG310736/c
LOCUS   AG310736   1359 bp   DNA   linear   GSS 02-JUN-2004
DEFINITION Mus musculus molossinus DNA, clone:MSMg01-093D20.TJ, genomic survey
            sequence.
ACCESSION   AG310736
VERSION   AG310736.1 GI:47883690
KEYWORDS   GSS.
SOURCE   Mus musculus molossinus
  ORGANISM Mus musculus molossinus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

AUTHORS   Ojopi,E.P.B., Paquola,A.C.M., Piazza,J.P., Nishiyama,M.Y. Jr.,
           Kitajima,J.P., Adamson,R.E., Ashton,P.D., Bonaldo,M.F.,
           Coulson,P.S., Dillon,G.P., Farias,L.P., Gregorio,S.P., Ho,P.L.,
           Leite,R.A., Malaquias,L.C.C., Marques,R.C.P., Miyasato,P.A.,
           Nascimento,A.L.T.O., Ohlweiler,F.P., Reis,E.M., Ribeiro,M.A.,
           Sa,R.G., Stukart,G.C., Soares,M.B., Gargioni,C., Kawano,T.,
           Rodrigues,V., Madeira,A.M.B.N., Wilson,R.A., Menck,C.F.M.,
           Setubal,J.C., Leite,L.C.C. and Dias-Neto,E.
TITLE   Transcriptome analysis of the acoelomate human parasite Schistosoma
           mansoni
JOURNAL   Nat. Genet. 35 (2), 148-157 (2003)
MEDLINE   22879926
PUBMED   12973350
COMMENT   Contact: Dr. Sergio Verjovski-Almeida
           Departamento de Bioquimica
           Instituto de Quimica - Universidade de Sao Paulo
           Av. Prof. Lineu Prestes 748 sala 1200, 05508-900 Sao Paulo - SP,
           Brasil
           Tel: +55-11-3091-2173
           Fax: +55-11-3091-2186
           Email: verjo@iq.usp.br
           This sequence was derived from the FAPESP Schistosoma mansoni EST
           Genome Project. All sequences in the project were assembled and
           annotated. This entry and all the assembled sequences can be seen
           in the following URL http://bioinfo.iq.usp.br/schisto/
           Plate: MM1-0013G-V027 row: 10 column: H.
FEATURES   Location/Qualifiers
     source   1..478
                     /organism="Schistosoma mansoni"
                     /mol_type="mRNA"
                     /db_xref="taxon:6183"
                     /clone="MM1-0013G-V027-H10.B"
                     /sex="male"
                     /dev_stage="adult"
                     /lab_host="Mesocricetus auratus"
                     /clone_lib="MM1-0013"
                     /note="Vector: SureClone"
ORIGIN

Query Match   2.8%;   Score 54.4; DB 6;   Length 478;
Best Local Similarity 52.2%;   Pred. No. 0.0056;
Matches 121; Conservative   0; Mismatches 111; Indels   0; Gaps   0;

Qy   616 TTCCAACAAGCTCTTGCTCAAAGAGACAATGTTAGTAAAAATGCCCCAGTTGCCGGTGGA 675
          || |||| |||| ||||| ||||| || ||||| |||| |||||| |||||||
Db   462 TTAAAACACTTCCGTACTTCAACAGCGAGCTCTGAGAAAACTTCGAGCGAACCTACTGCA 403

Qy   676 AGAGGTGAAGAACAACGGATGATGAATCGAGTGGACCAAAGAATGCAACAAAGAGAA 735
          |||| ||||||| ||| ||| || |||||||||| | | |||| ||||||||
Db   402 CCAGCTGATGAACAACAGACAACAGACACGCCCGAAACAAACTGAACAAAGTGTTGACGAAACCTCA 343

Qy   736 CTTCAAGAGGAAGATGAAGATGATGAATCGTTGAGGACGAGGATGTACCCAGAAGAAGA 795
          | |||| ||||| ||| ||| |||| ||| |||||| || |||| ||||||||||
Db   342 CCTAAAGATGAAACTAAAGAAAAGATCAGCCTGTTGAACCAAATGAAAAAAAGAAATC 283

Qy   796 AGTTCGGATGGAGAACCACAAAGTGAAGCAGCAGAGCATCAGAGAAGAAGAGATTTAG 847
          ||||| |||||||||||||||||| | |||||||||| || || |||||||| |||
Db   282 CAAGAAATGTTGAAGTTGTAGGAGGTAACAGAGCAACCACAACCATCTGTAG 231

RESULT 6
CD185097
LOCUS   CD185097   534 bp   mRNA   linear   EST 14-SEP-2003
DEFINITION MS1-0051T-L290-A06-U.B MS1-0051 Schistosoma mansoni cDNA clone
            MS1-0051T-L290-A06.B, mRNA sequence.
ACCESSION   CD185097
VERSION   CD185097.1 GI:34715241
KEYWORDS   EST.
SOURCE   Schistosoma mansoni
  ORGANISM Schistosoma mansoni
            Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea;
            Strigeidida; Schistosomatoidea; Schistosomatidae; Schistosoma.
REFERENCE   1 (bases 1 to 534)

```
REFERENCE   1
AUTHORS     Hattori,M., Toyoda,A., Noguchi,H., Kojima,T. and Sakaki,Y.
TITLE       BAC end Sequences of Library MSMg01
JOURNAL     Unpublished
REFERENCE   2  (bases 1 to 1359)
AUTHORS     Hattori,M., Toyoda,A., Noguchi,H., Kojima,T. and Sakaki,Y.
TITLE       Direct Submission
JOURNAL     Submitted (17-NOV-2003) Masahira Hattori, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou.Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            (E-mail:hattori@gsc.riken.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT     Clones are derived from the mouse BAC library MSMg01. For BAC
            library availability, please contact Kuniya Abe (abe@rtc.riken.jp).
            The Institute of Physical and Chemical Research (RIKEN) 3-1-1
            Koyadai, Tsukuba, 305-0074 Japan
            phone: 81-298-36-9189, fax: 81-298-36-9199
            e-mail: abe@rtc.riken.jp
PRIMERS
            Sequencing : TJ
LIBRARY
            Vector     : pBACe3.6
            R.Site 1   : EcoRI
            R.Site 2   : EcoRI.
FEATURES    Location/Qualifiers
  source    1..1359
            /organism="Mus musculus molossinus"
            /mol_type="genomic DNA"
            /sub_species="molossinus"
            /db_xref="taxon:57486"
            /clone="MSMg01-093D20.TJ"
            /sex="male"
            /tissue_type="mixture of kidney and spleen"
            /clone_lib="MSMg01 Mouse Male BAC Library"
ORIGIN

Query Match         2.8%;  Score 53.8;  DB 9;  Length 1359;
Best Local Similarity  45.8%;  Pred. No. 0.011;
Matches  299;  Conservative  0;  Mismatches 347;  Indels  7;  Gaps  3;

Qy    232  AAGACTCAATTAGAAGTTCTAGAGAAAGTACATCCTGATCAATTCGATAAGTACAAAAAG  291
           |||||    |||||| ||||||| ||||||    ||||||||||||||||
Db   1185  AAGAAAAGAAGAGATGATAGAGAAATGAAATGAAAAATCGATAAGTAGAATATG  1126

Qy    292  CTAAAAGTTGATGATTTGGCAGCTGATGCAGTTATGCAACAGGCGGAGATGGCAAAATTA  351
           |||  ||| ||||   ||  |||||||    ||||||||||||||||||
Db   1125  AAAAATTATGAAAAGAAGAGAGAAAGAAGAAGAAATAGAGAGAGAAAAATGAGAATAGAA  1066

Qy    352  CAGCCTAAATCAGGAAATGCATTTATCGATATGTTGAATG---GAAATGGAATCCCAATT  408
           |||  ||||  ||||||||||||||||||   |||||||    ||||||||||||
Db   1065  GAGAGAAAAAGAGGAAAAAGAAGAAAGAAATAAAAAAAGAATAAGGAAGAGAAGAAAGA  1006

Qy    409  GGTAGCAGTATTCGTGGATTAGAAGAATGCTATCCGAACGCAGAGAGATGTGAAAATACG  468
           ||||||| |||||  |    ||||||||||||||| |||||||||||||  |||||||
Db   1005  GAAAAAGAATAGGAAAAAAGAAAAGAAAGAGAGAAAATAAAATAAGAA  946

Qy    469  GATCCGTCCGAACAGATTGCCAAAGCCGTAATGGACAAATTTCAAACACAAATTCTCCCA  528
           |||| ||||||||||  ||||||||||||||||| ||
Db    945  AAAGAGAAAGAA-AGATAAAAGAATAAATAATGAAAAAATAATTATAGAGAAGAAGAGA  887

Qy    529  GGACTCGTTGCAAATATGATCCGTGGCAAGAACCCCTTTAAAATGCCTCAACAAATGAGA  588
           ||||  |||||||| ||||   ||||||||||||||||| ||
Db    886  AAAAATAT---AAAATAGAGAAAAGAGATAAAAAATGAAATGTAAGAAAAAAAAAAAAA  830

Qy    589  AAAGCTCAAGCTGCTCCATCGTCAGTGTTCCAACAAGCTCTTGCTCAAAGAGCCAATGTTA  648
           ||||        |||| ||||||||||||||| |||| ||||||||
Db    829  TAAATATGAAAAAGAGAAGAAGAAGAAGGAAGAAAATTATAAAAAAAAAAAAAGAAGAAA  770

Qy    649  GGTAAAAAATGCCCCAGTTGCCCGGTGAAGAGGTGAAGAACAACGGATGATGAATCGA  708
           ||  ||||  ||||
Db    769  GAAAGAAAAAAATAAGTAAGAAGAGAGGAAGAGGGAAAGAGAAAAAGATAAAAGAAGAGA  710
```

```
Qy    709  GTGGACCAAAGAATGCAACAAAGAGAACTTCAAGAGGAAGATGAAGAGATGAGATGATGATCTT  768
           ||  ||||    |||||| ||||||||   |||||| ||||||| ||| ||| |||
Db    709  GTAGATGAAAGAAGAAAGTAAGAAGGCAAGAGAAAAGAGATAAGAAGAAAGAAAAA  650

Qy    769  GAGGACGAGGATGTACCCAGAAGAAGAGTTCGGATGGAGAACCACAAAGTGAAGCAGAG  828
           ||||||||||||||||||||||||||||||| ||||||||||  |||| |||||
Db    649  GAAGACAATAAAGAAAAGAAAAAAAAAGAAGAGAAGAAAGAGAAAAAGGGAACAA  590

Qy    829  CATCAGAGAGAGATTTAGCCAGAGAGATTGAAAAGTAGTCCTAGATTAAAAGA  881
           |||||| |||||| |||| |||||  ||||||||
Db    589  AAAAAAAGAGAGATACAAAAAGAAATAGAAACAAAGAAAAATGAAGAAA  537

RESULT 8
LOCUS       BU418103
BU418103    603959413F1 CSEQRBN09 Gallus gallus cDNA clone ChEST93214 5', mRNA
DEFINITION  622 bp    mRNA        linear  EST 29-NOV-2002
            sequence.
ACCESSION   BU418103
VERSION     BU418103.1  GI:25910774
KEYWORDS    EST.
SOURCE      Gallus gallus (chicken)
  ORGANISM  Gallus gallus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
            Phasianinae; Gallus.
REFERENCE   1  (bases 1 to 622)
AUTHORS     Boardman,P.E., Sanz-Ezquerro,J., Overton,I.M., Burt,D.W., Bosch,E.,
            Fong,W.T., Tickle,C., Brown,W.R.A., Wilson,S.A. and Hubbard,S.J.
TITLE       A Comprehensive Collection of Chicken cDNAs
JOURNAL     Curr. Biol. 12 (22), 1965-1969 (2002)
MEDLINE     22335534
PUBMED      12445392
COMMENT     Contact: Simon Hubbard
            Department of Biomolecular Sciences
            University of Manchester Institute of Science and Technology
            (UMIST)
            PO Box 88, Manchester, M60 1QD, UK
            Tel: 01612008930
            Fax: 01612360409
            Email: Simon.Hubbard@umist.ac.uk.
FEATURES    Location/Qualifiers
  source    1..622
            /organism="Gallus gallus"
            /mol_type="mRNA"
            /strain="Layer and broiler"
            /db_xref="taxon:9031"
            /clone="ChEST93214"
            /sex="Male and female"
            /tissue_type="Chondrocytes isolated from growth plate
            cartilage"
            /dev_stage="adult"
            /lab_host="DH10B"
            /clone_lib="CSEQRBN09"
            /note="Vector: pBluescript II KS(+); Site 1: EcoRI;
            Site 2: NotI; This normalized library was constructed from
            1 million independent clones. cDNA synthesis was initiated
            using an oligo(dT) primer, using methylated C in the first
            strand synthesis reaction. Following this first strand
            reaction, double-stranded cDNA was blunted, ligated to
            NotI adapters, digested with EcoRI, size-selected, and
            cloned into the NotI and EcoRI compatible sites of a
            custom modified MCS of the pBluescript (KS+) vector. The
            library was normalized in 2 rounds using conditions
            adapted from Soares et al., PNAS (1994) 91: 9228-9232 and
            Bonaldo et al., Genome Research 6 (1996): 791, except that
            a significantly longer reannealing hybridization was
            used."
ORIGIN

Query Match         2.8%;  Score 53.6;  DB 5;  Length 622;
Best Local Similarity  51.2%;  Pred. No. 0.0098;
Matches  125;  Conservative  0;  Mismatches 119;  Indels  0;  Gaps  0;
```

ligated to the blunt ends in high molar excess. The
adaptored DNA was purified and size-selected for a 9.5 to
10.5 kb range using preparative agarose gel
electrophoresis. Vector DNA was prepared from a derivative
of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
inducible derivative of plasmid R1. The vector was ligated
with adaptors complementary to the insert adaptors and
purified. The sheared, adaptored mouse DNA was annealed to
adaptored vector DNA, and transformed into
chemically-competent E. coli XL10-Gold (Stratagene) cells
and selected for ampicillin resistance."

ORIGIN

Query Match   2.7%;   Score 53; DB 8; Length 488;
Best Local Similarity   50.2%;   Pred. No. 0.013;
Matches 131; Conservative 0; Mismatches 130; Indels 0; Gaps 0;

```
Qy  674  GAAGAGGTGAAGAACAACGATGATGATGAATCGAGTGACCAAAGAATGCAACAAAGAG  733
         ||||| || |||||||                    ||       ||||||||||||
Db   42  GAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAG  101

Qy  734  AACTTCAAGAGAGAAGATGAAGATGATGAAGATGATCTTGAGGACGAGGATGTACCCAGAGAA  793
             ||||||||||||||| |||| || |||||| || ||       |||||||||||||||||
Db  102  AAGAGAGAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAA  161

Qy  794  GAAGTTCGGATGGAGAACCCACAAAGTGAAGCAGAGCATCAGAGAGAGATTTAGCCAGGA  853
         ||||  | |||         ||| |||||  |||| |||||||||||||||    ||||||
Db  162  GGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGA  221

Qy  854  GATTGAAAAGTAGTCCTAGATTAAAAGAGCTTTTACAGAATGCGGAAGTTCAATCATTGC  913
             ||||| | |||||||                                        |||
Db  222  GAAGAAGAAGAAGAAGAAGAAGAAGAAGTAGTAGTAGTAGTGGTGTGTCTCGTTGTTC  281

Qy  914  TCTCTTACCAACGAATGAGGG  934
             |||  |||||||    |||
Db  282  TTCCTTGCCAGCTTATTCTGG  302
```

RESULT 10
CNS0207H
LOCUS   CNS0207H   1196 bp   DNA   linear   GSS 01-SEP-2000
DEFINITION   Tetraodon nigroviridis genome survey sequence T7 end of clone
             152P14 of library G from Tetraodon nigroviridis, genomic survey
             sequence.
ACCESSION   AL206342
VERSION   AL206342.1   GI:7865161
KEYWORDS   GSS; genome survey sequence.
SOURCE   Tetraodon nigroviridis
  ORGANISM   Tetraodon nigroviridis
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
             Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
             Tetraodontoidea; Tetraodontidae; Tetraodon.
REFERENCE   1
  AUTHORS   Roest Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
             Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
             Saurin,W. and Weissenbach,J.
  TITLE   Estimate of human gene number provided by genome-wide analysis
             using Tetraodon nigroviridis DNA sequence
  JOURNAL   Nat. Genet. 25 (2), 235-238 (2000)
  MEDLINE   20296633
  PUBMED   10835645
REFERENCE   2
  AUTHORS   Roest Crollius,H., Jaillon,O., Dasilva,C., Dasilva,C., Ozouf-Costaz,C.,
             Fizames,C., Fischer,C., Bouneau,L., Billault,A., Quetier,F.,
             Saurin,W., Bernot,A. and Weissenbach,J.
  TITLE   Characterization and repeat analysis of the compact genome of the
             freshwater pufferfish Tetraodon nigroviridis
  JOURNAL   Genome Res. 10 (7), 939-949 (2000)
  MEDLINE   20359837
  PUBMED   10899143
REFERENCE   3 (bases 1 to 1196)
  AUTHORS   Genoscope.

```
Qy  663  AGTTGCCGGTGGAAGAGGTGAAGAACAACGATGATGATGAATCGAGTGACCAAAGAAT  722
         |||||  ||||||| |||||||  ||          ||   || ||    |  |||||
Db  120  GAAGATGAAGATGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGA  179

Qy  723  GCAACAAAGAGAACTTCAAGAGAGAAGATGAAGATGATGATGAAGATGATCTTGAGGACGAGGATGT  782
         |    || |  ||  ||  || ||||||||| || |||| |||||| |||||| ||   ||||||||
Db  180  AGATGAAGAAGAAGAAGAAGAGATGAAGATGAAGATGAAGATGAAGATGAAGAAGATGA  239

Qy  783  ACCCAGAGAAGAAGTTCGGATGGAGAACCCACAAAGTGAAGCAGAGCATCAGAGAGAGA  842
         |   |||||||||  |  |||    |||    |||| | ||| ||||| |||||||||
Db  240  AGATGAAGATGAAGATGAAGATGGGAAGAAGAAGAAGAAGAGGAGGAGATGAGCAGAGATGAGC  299

Qy  843  TTTAGCCAGGAGATTGAAAAGTAGTCCTAGATTAAAAGAGCTTTTACAGAATGCGGAAGT  902
         | |||   ||||| ||||| |                  ||||||| || || |||
Db  300  AAGACTGAAGAAGAAGAAGACGAATGCAGATGACCATGCAGCCATGAAGAAGA  359

Qy  903  TCAA  906
            ||
Db  360  AGAA  363
```

RESULT 9
AZ379487
LOCUS   AZ379487   488 bp   DNA   linear   GSS 02-OCT-2000
DEFINITION   1M0134N11R Mouse 10kb plasmid UUGC1M library Mus musculus genomic
             clone UUGC1M0134N11 R, genomic survey sequence.
ACCESSION   AZ379487
VERSION   AZ379487.1   GI:10493187
KEYWORDS   GSS.
SOURCE   Mus musculus (house mouse)
  ORGANISM   Mus musculus
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 488)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
             Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T.,
             Reilly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von
             Niederhausern,A. and Wright,D., Weiss,R.
  TITLE   Mouse whole genome scaffolding with paired end reads from 10kb
             plasmid inserts
  JOURNAL   Unpublished (2000)
COMMENT   Contact: Robert B. Weiss
             University of Utah Genome Center
             University of Utah
             Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
             84112, USA
             Tel: 801 585 5606
             Fax: 801 585 7177
             Email: ddunn@genetics.utah.edu
             Insert Length: 10000   Std Error: 0.00
             Plate: 0134 row: N column: 11
             Seq primer: CACACAGGAAACAGCTATGACC
             Class: plasmid ends
             High quality sequence stop: 488.
FEATURES   Location/Qualifiers
     source   1..488
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UUGC1M0134N11"
                     /sex="Male"
                     /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                     /clone_lib="Mouse 10kb plasmid UUGC1M library"
                     /note="Vector: pWD42nv; Purified genomic DNA from M.
                     musculus C57BL/6J (male) was obtained from the Jackson
                     Laboratory Mouse DNA Resource
                     (http://www.jax.org/resources/documents/dnares/). The DNA
                     was hydrodynamically sheared by repeated passage through a
                     0.005 inch orifice at constant velocity. The sheared DNA
                     was blunt end-repaired with T4 DNA polymerase and T4
                     polynucleotide kinase. Adaptor oligonucleotides were

Left column

TITLE   Direct Submission
JOURNAL   Submitted (12-APR-2000) Genoscope - Centre National de Sequencage :
          BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
          - Web : www.genoscope.cns.fr)
COMMENT   This sequence is a single read and was generated as part of a large
          scale clone-end sequencing project of the Tetraodon nigroviridis
          genome. For more information, please take a look at
          http://www.genoscope.cns.fr/Tetraodon.
FEATURES Location/Qualifiers
  source 1..1196
          /organism="Tetraodon nigroviridis"
          /mol_type="genomic DNA"
          /db_xref="taxon:99883"
          /clone="152P14"
          /clone_lib="G"
          /note="Genoscope sequence ID : C0AG152DH07LP1-end : T7"
ORIGIN

Query Match   2.7%; Score 53; DB 9; Length 1196;
Best Local Similarity 52.0%; Pred. No. 0.017;
Matches 119; Conservative 0; Mismatches 110; Indels 0; Gaps 0;

```
Qy  653 AAAATGCCCAGTTGCCGGTGGAAGAGTGAAGAACACGGATGATGATGAATCGAGTGG 712
        | ||||| |||  ||  ||||||||||||||||||||||||| ||||||
Db  120 AGAACCGCGCAGGGGAAGAGGAAGAAGAGAAGAAGAAGAAGAAGAAGAAGAAGAAGA 179

Qy  713 ACCAAAGAATGCAACAAAGAGACTTCAGAGAGAAGAGAAGATGAAGATGATGATCTTGAGG 772
        |  |||||  |  |||
Db  180 AGAGGAAGAAGAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGA 239

Qy  773 ACGAGGATGTACCCAGAAGAAGAAGTTCGGATGGAGAACCACCAAAGTGAAGCAGAGCATC 832
        |  ||| |||                  |||||||  ||||  ||
Db  240 AGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAGA 299

Qy  833 AGAGAAGAGATTTAGCCAGGACAGATTGAAAAGTAGTCCTAGATTAAAGA 881
        ||||||||| |||||
Db  300 AGAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAGAGAGA 348
```

RESULT 11
CR089470/c
LOCUS   CR089470   610 bp   DNA   linear   GSS 05-JUL-2004
DEFINITION Reverse strand read from insert in 5'HPRT insertion targeting and
            chromosome engineering clone MHPN28i17, genomic survey sequence.
ACCESSION   CR089470
VERSION   CR089470.1 GI:49823062
KEYWORDS   GSS; genome survey sequence; MICER.
SOURCE   Mus musculus (house mouse)
  ORGANISM Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 610)
  AUTHORS   Adams,D.J., Biggs,P.J., Cox,A.V., Davies,R.M., van der Weyden,L.,
            Jonkers,J., Smith,J., Plumb,R.W., Taylor,R.G., Nishijima,I., Yu,Y.,
            Rogers,J. and Bradley,A.
  TITLE   Direct Submission
  JOURNAL   Submitted (20-FEB-2004) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. http://www.sanger.ac.uk/MICER
FEATURES   Location/Qualifiers
  source   1..610
            /organism="Mus musculus"
            /mol_type="genomic DNA"
            /db_xref="taxon:10090"
            /clone="MHPN28i17"
            /clone_lib="MHPN"
ORIGIN

Query Match   2.6%; Score 50.8; DB 9; Length 610;
Best Local Similarity 55.7%; Pred. No. 0.054;
Matches 97; Conservative 0; Mismatches 77; Indels 0; Gaps 0;

```
Qy  682 GAAGAACAACGATGATGATGATGAATCGAGTGGACCAAGAATGCAACAAAGAGAAGAACTTCAA 741
        ||||| |||  ||  ||
```

Right column

```
Db  497 GAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAA 438
Qy  742 GAGGAAGATGAAGATGATGATGATCTTGAGGACGACGATGTACCCAGAAGAAGAAGAAGAGTTCG 801
        || |||||| ||| ||  |
Db  437 GAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAAAAGAAGAAGAAGAAGAAGAAGAAGAAGA 378
Qy  802 GATGGAGAACCACCAAAGTGAAGCAGAGCATCAGAGAAGAGATTAGCAGCTAAAGGA 855
        |  ||| ||| |||    ||
Db  377 AAAGAAAAGAAAACAAAGAAAAGAAGAAGAAGAGAAGAAGAGCTAAAGGA 324
```

RESULT 12
CR214149/c
LOCUS   CR214149   838 bp   DNA   linear   GSS 06-JUL-2004
DEFINITION Reverse strand read from insert in 5'HPRT insertion targeting and
            chromosome engineering clone MHPN328h07, genomic survey sequence.
ACCESSION   CR214149
VERSION   CR214149.1 GI:49992998
KEYWORDS   GSS; genome survey sequence; MICER.
SOURCE   Mus musculus (house mouse)
  ORGANISM Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 838)
  AUTHORS   Adams,D.J., Biggs,P.J., Cox,A.V., Davies,R.M., van der Weyden,L.,
            Jonkers,J., Smith,J., Plumb,R.W., Taylor,R.G., Nishijima,I., Yu,Y.,
            Rogers,J. and Bradley,A.
  TITLE   Direct Submission
  JOURNAL   Submitted (20-FEB-2004) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. http://www.sanger.ac.uk/MICER
FEATURES   Location/Qualifiers
  source   1..838
            /organism="Mus musculus"
            /mol_type="genomic DNA"
            /db_xref="taxon:10090"
            /clone="MHPN328h07"
            /clone_lib="MHPN"
ORIGIN

Query Match   2.6%; Score 50.8; DB 9; Length 838;
Best Local Similarity 55.7%; Pred. No. 0.059;
Matches 97; Conservative 0; Mismatches 77; Indels 0; Gaps 0;

```
Qy  682 GAAGAACAACGATGATGATGATGAATCGAGTGGACCAAGAATGCAACAAAGAGAAGAACTTCAA 741
        ||||| |||  ||  ||
Db  497 GAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAA 438
Qy  742 GAGGAAGATGAAGATGATGATGATCTTGAGGACGACGATGTACCCAGAAGAAGAAGAAGAGTTCG 801
        || |||||| ||| ||  |
Db  437 GAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAAAAGAAGAAGAAGAAGAAGAAGAAGAAGA 378
Qy  802 GATGGAGAACCACCAAAGTGAAGCAGAGCATCAGAGAAGAGATTAGCAGCTAAAGGA 855
        |  ||| ||| |||    ||
Db  377 AAAGAAAAGAAAACAAAGAAAAGAAGAAGAAGAGAAGAAGAGCTAAAGGA 324
```

RESULT 13
CNS0037Q
LOCUS   CNS0037Q   1101 bp   DNA   linear   GSS 03-JUN-1999
DEFINITION Drosophila melanogaster genome survey sequence TET3 end of BAC #
            BACR08K14 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL064465
VERSION   AL064465.1 GI:4941722
KEYWORDS   GSS.
SOURCE   Drosophila melanogaster (fruit fly)
  ORGANISM Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1 (bases 1 to 1101)
  AUTHORS   Genoscope.
  TITLE   Direct Submission

JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
           BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
           - Web : www.genoscope.cns.fr)
COMMENT   Determination of this BAC-end sequence was carried out as part of a
           collaboration with the Berkeley Drosophila Genome Project (BDGP).
           The BDGP is constructing a physical map of the Drosophila
           melanogaster genome using these BACs. For further information
           please see http://www.fruitfly.org The BDGP Drosophila
           melanogaster BAC library was prepared by Kazutoyo Osoegawa and
           Aaron Mammoser in Pieter de Jong's laboratory in the Department of
           Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
           NY. The library is named RPCI-98 and was constructed by partial
           EcoRI digestion of Drosophila DNA provided by the BDGP from the
           isogenic strain y2; cn bw sp, the same strain used for the BDGP's
           P1 and EST libraries. A more detailed description of the library
           and how to order individual BAC clones, the entire library, or
           filters for hybridization from the BACPAC Resource Center can be
           found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES   Location/Qualifiers
     source   1..1101
                   /organism="Drosophila melanogaster"
                   /mol_type="genomic DNA"
                   /db_xref="taxon:7227"
                   /clone="BACR08K14"
                   /clone_lib="RPCI-98"
                   /note="end : TET3"
ORIGIN

Query Match   2.6%;   Score 50.8;   DB 9;   Length 1101;
Best Local Similarity 16.5%;   Pred. No. 0.064;
Matches   81; Conservative 202; Mismatches 205; Indels   2; Gaps   2;

Qy   578 AACAAATGAGAAAAGTCAAGCTCCATCGTCCATCGTCAGTGTTCCAACAAGCTCTTGCTCAA 637
             ||::| ::   : |:|:|:|: :|:|:|:
Db   564 AAACMAAMANAACAGCCCTCBYHHCBCYAYNYHHBAGHTKCTARHRYDSAYDTRMRTGAG 623

Qy   638 GAGCAATGTTAGGTAAA-ATGCCCCAGTTGCCGGTGGAGAGGTGAGACAACGACGATG 696
         : ::|::::::::| :: ::| :|::| : ::::|
Db   624 DGGAKRYANKDGRKDDVRKGNKMWASAVANGAADNMTRMRTACANGWAMKTWHAMKKKDT 683

Qy   697 ATGATGAATCGAGTGGACCAAAGAATGCAACAAAGAGACTTCAAGAGAGAAGATGAAGAT 756
         :: ::|:|:|| ::|: ::|:| :::| : :||::|
Db   684 GKRNNKAKAKWGRVWBDNVKAATTDTMTKTRMAVAMARAGAAMARTAMRARAAKAARAK 743

Qy   757 GATGATGATCTTGAGGACGAGGAGTGTACCCAGAAGAAGAAGTTCGGATGGAGAACCACAA 816
         ::|:||::|: : :::: :| :|::|:|:| :|:|:|:|
Db   744 KANKKKGTKATKDKMGMKGMKGMGWGKHKKKKKDMKMDGAKATAWAMTRAGAVGVKARRAMKATRW 803

Qy   817 AGTGAAGCAGAGCATCAGAGAAGAGATTTAGCCACGAGATTGAAAAGTAGTCCTAGATTA 876
         || ::|:::::| :   : | ||:: : |:|: |:
Db   804 AGAARRVKAKRGAMRVMKADAM-RRMRHWARWMANGAGHNTVNMBVBVNMGNGMKSKGMGN 862

Qy   877 AAAGAGCTTTTACAGAATGCGGAAGTTCAATCATTGCTCTCTTACCAACGAATGAGGGAT 936
             :| :::::|:|:|:::| :| ::||:|
Db   863 TKNKVNSVKTVKBHKDHKTTTTTNTNMNMTHTTHVMBMBMGMGMRMDMGMGRGVVGHRV 922

Qy   937 TCTCCACTGAGCAAGCGAAGCCCTTTGGCTATGAACGATGAGGAGAAGTGCATTCCGC 996
         :: :::|::::: ::::| ::|:::|:: : :: |: |
Db   923 GKMNGMNRMVMRMVNMVKVKAVAVMGMMVMVMMMVMGAMVMVMMNMGMGYSGSGSKVSCK 982

Qy   997 GCAATGGAGGCTCGTGCAAAACTAGATCAAAAATCTCAACTTGTCTGCTCCATGGT 1056
         :::|:::| : ::: ::|:|:||: :|::::|: :| :|
Db   983 GSTMGKKNGVGMVCVBNMKVXMGNKMKMGMNVTVMGMKMBMBVNMKVNMKMKMGRTNKR 1042

Qy 1057 TTTGGAGAGT 1066
         |:|:|:|::
Db 1043 TMKMGHTHGW 1052

RESULT 14
CL111125/c
LOCUS   CL111125   1133 bp   DNA   linear   GSS 05-JAN-2004
DEFINITION ISB1-54M20_T7.1 ISB1 Xenopus tropicalis genomic clone ISB1-54M20,

ACCESSION   CL111125
VERSION   CL111125.1 GI:40504760
KEYWORDS   GSS.
SOURCE   Xenopus tropicalis (western clawed frog)
  ORGANISM Xenopus tropicalis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
            Xenopodinae; Xenopus; Silurana.
REFERENCE   1 (bases 1 to 1133)
  AUTHORS   Kremitzki,C., Carter,J., McPherson,J., Warren,W., Graves,T.,
            Mardis,E. and Wilson,R.
  TITLE   A physical map of the xenopus tropicalis genome
  JOURNAL   Unpublished (2003)
COMMENT   Contact: Richard K Wilson
            Genome Sequencing Center
            Washington University School of Medicine
            Email: submissions@watson.wustl.edu
            Insert Length: 75000   Std Error: 0.00
            Seq primer: T7 TAATACGACTCACTATAGGG
            Class: BAC ends
            High quality sequence start: 125
            High quality sequence stop: 381.
FEATURES   Location/Qualifiers
     source   1..1133
                   /organism="Xenopus tropicalis"
                   /mol_type="genomic DNA"
                   /db_xref="taxon:8364"
                   /clone="ISB1-54M20"
                   /clone_lib="ISB1"
                   /note="Vector: pBeloBAC11; ISB1-1 Xenopus tropicalis BAC
                   Library Segment 1"
ORIGIN

Query Match   25.6%;   Score 50.8;   DB 9;   Length 1133;
Best Local Similarity   2.6%;   Pred. No. 0.064;
Matches 196; Conservative 0; Mismatches 571; Indels 0; Gaps   0;

Qy   197 AAAAAGGTCAAATGGAAATGGCAAGGAGCATTYAAGACTCAATTAGAAGTTCTAGAGA 256
         |||||||| ::: : :|| ::::::::
Db 1133 AAAAAANNANANANAAANAAAAAAAAAANAAANAAAGTCAAAAAGCTAAAAGTTGGCAGCTG 1074

Qy   257 AAGTACATCCTGATCAATTCGATAGTACAAAAAGTCAAAATGAAGANAAANNANAANNA 316
         |||||
Db 1073 AANAAAAAAAAAAAAANAAAAANAAAAAAAAANAAAANAANNANAAANNA 1014

Qy   317 ATGCAGTTATGCAACAGGCGGAGATGGCAAAATTACAGCCTAAATCAGGAGAATGCATTA 376
         |||||
Db 1013 AAANANNANAAANNANNNNAAAAANNAAANNAAAAAANAAAAANNAAANNANNN 954

Qy   377 TCGATATGTTGAATGAAATGGAAATGCCCAATTGGTAGCAGTATTCGTGGATTAGAAGATG 436
         |||||
Db   953 ANAAAAAAAANAAANAANNANNAANNAAAAAAANAAAAAAANAAAAAN 894

Qy   437 CTATCCGAACGCAGAGAGATATGGAAAATACGGATCCGTCCGAACAGATTGCCAAAGCCG 496
         |||||
Db   893 NNAAANAAAAAAANAAAAAANAANNANNANNAAAAAAAANAAAAAANNAN 834

Qy   497 TAATGACAAATTTCAAACACAAATTCTCCAGGACTCGTTGCAAATATGCTCGTGGCA 556
         |||||
Db   833 NNNNNNNNNNAAAAAAAANNANNAAANNNNNNANAAANNAAAANNNNNAANNA 774

Qy   557 AGAACCCCTTTAAAATGCCTCCAACAAATGAGAAAAGCTCAGCGTCTCCATCGTCAGTGT 616
         |||||
Db   773 AAAAANNNANNANNAAAAANNNNNNAAAANNNNNAAANNNNAAAANNNNAAAA 714

Qy   617 TCCAACAAGCTCTTGCTCAAGAGCAATGTTAGGTAAAAATGCCCCAGTTGCCGGTGGAA 676
         |||||
Db   713 ANNAAANANNANNNNAANNNNANANNNNNNANNAANNNNANNAAAAAAANANANANAAAA 654

Qy   677 GAGGTGAAGAACAACCGATGATGAATCGAGTGGCAAGAATGCAACAAAGAGAAC 736
         |||||
Db   653 ANNNANAAAAGANANNANNNNNNNNNNNNANNAAANAAAAAANANAANNAAANAAAAANANAAA 594

Matches 110; Conservative 0; Mismatches 99; Indels 0; Gaps 0;

```
Qy    673 GGAAGAGGTGAAGAACCACCGGATCGTGATGATGATCTGAGTCGAGTGGACCCAAAGAATGCAACAAAGA 732
              | || | |||||||||||  | ||  ||||||  |||||| ||| |||||  |||
Db   1194 GAAGAGAAGAAGAAGAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAGAAGAAGA 1135

Qy    733 GAACTTCAAGAGAAGATGAAGATGATGATCTTGAGGACGAGGATGTGACCCAGAAGA 792
              |  |||| ||||||||||||  |||  ||||||| ||  ||| |||||||
Db   1134 AGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAGAGAGAGA 1075

Qy    793 AGAAGTTCGGATGAGAACCACAAAGTGAAGCAGCATCAGAAGAAGAGATTTAGCCAGG 852
              ||||  |||| |||||||||||| |||  ||||||  |||||||||||| |||||
Db   1074 AGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAG 1015

Qy    853 AGATTGAAAAGTAGTCCTTAGATTAAAAGA 881
              ||||| |||  |||| |||||
Db   1014 AGAAGAAGAAGAAGAAGAAGAAGAAGAGAAGAAGA 986
```

```
Qy    737 TTCAAGAGGAAGATGAAGATGATGATGATCTCTTGAGGACGTGACCCAGAAGAAGAA 796
              |  |   |     |  || || |||||
Db    593 ANNNNAANNNAAAAAANNNANAAAAANNNNNAAANNNNAAAANNNNNAAAAAAAAAAN 534

Qy    797 GTTCGGATGGAGAACCACAAAGTGAAGCAGCATCAGAGAGAGATTTAGCCAGGAGAT 856
              |  | |   |   | |   |  ||
Db    533 NNANNNNNAAAAAANAAAANNNNNAAAAAAAAAAANNNNNNAANNNNNNAAAANNNN 474

Qy    857 TGAAAAGTAGTCCTAGATTAAAAGAGCTTTTACAGAATGCGAAGTTCAATCATTGCTCT 916
              |   |  |           |  | || 
Db    473 NNNNNNNANNNNNANANNNNNNAAAAAAAANNNNNNNNNNANNNNNNNN 414

Qy    917 CTTACCAACGAATGAGGGATTCTCCACTGAGCAAGCGAAGGCCTTTG 963
              |  | |||||    | |    |
Db    413 NNNNAAAAAAAANNNNNNNNNNNNNNNCCCCCAAGTGACTTAG 367
```

Search completed: July 15, 2005, 14:57:35
Job time : 6512 secs

RESULT 15
CNS02O3O/c

```
LOCUS       CNS02O3O        1201 bp    DNA        linear   GSS 01-SEP-2000
DEFINITION  Tetraodon nigroviridis genome survey sequence T7 end of clone
            152B16 of library G from Tetraodon nigroviridis, genomic survey
            sequence.
ACCESSION   AL206205
VERSION     AL206205.1  GI:7865024
KEYWORDS    GSS; genome survey sequence.
SOURCE      Tetraodon nigroviridis
  ORGANISM  Tetraodon nigroviridis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
            Tetradontoidea; Tetraodontidae; Tetraodon.
REFERENCE   1
  AUTHORS   Roest Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
            Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
            Saurin,W. and Weissenbach,J.
  TITLE     Estimate of human gene number provided by genome-wide analysis
            using Tetraodon nigroviridis DNA sequence
  JOURNAL   Nat. Genet. 25 (2), 235-238 (2000)
  MEDLINE   20296633
   PUBMED   10835645
REFERENCE   2
  AUTHORS   Roest Crollius,H., Jaillon,O., Dasilva,C., Ozouf-Costaz,C.,
            Fizames,C., Fischer,C., Bouneau,L., Billault,A., Quetier,F.,
            Saurin,W., Bernot,A. and Weissenbach,J.
  TITLE     Characterization and repeat analysis of the compact genome of the
            freshwater pufferfish Tetraodon nigroviridis
  JOURNAL   Genome Res. 10 (7), 939-949 (2000)
  MEDLINE   20359837
   PUBMED   10899143
REFERENCE   3  (bases 1 to 1201)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-APR-2000) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     This sequence is a single read and was generated as part of a large
            scale clone-end sequencing project of the Tetraodon nigroviridis
            genome. For more information, please take a look at
            http://www.genoscope.cns.fr/Tetraodon.
FEATURES             Location/Qualifiers
     source          1..1201
                     /organism="Tetraodon nigroviridis"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:99883"
                     /clone="152B16"
                     /clone_lib="G"
                     /note="Genoscope sequence ID : C0AG152DA08LP1-end : T7"
ORIGIN
```

Query Match   2.6%;   Score 50.6; DB 9; Length 1201;
Best Local Similarity 52.6%; Pred. No. 0.074;

```
              GenCore version 5.1.6
       Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on:   July 18, 2005, 12:25:30 ; Search time 7749 Seconds
                                    (without alignments)
                              4020.739 Million cell updates/sec

Title:          US-10-736-868-2
Perfect score:  3291
Sequence:    1 MILFLFLLLGFCIAPLSA.........SKTKRFVGGNGAFDMPALGL 643

Scoring table:  BLOSUM62
          Xgapop 10.0 , Xgapext  0.5
          Ygapop 10.0 , Ygapext  0.5
          Fgapop  6.0 , Fgapext  7.0
          Delop   6.0 , Delext   7.0

Searched:  4708233 seqs, 2422607955 residues

Total number of hits satisfying chosen parameters:     9416466

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US10736868/runat_14072005_105509_2278/app_query.fasta_1.839
-DB=GenEmbl -QFMT=fastap -SUFFIX=rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10736868 @CGN_1_1_4545 @runat_14072005_105509_2278 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :   GenEmbl:*
         1:  gb_ba:*
         2:  gb_htg:*
         3:  gb_in:*
         4:  gb_om:*
         5:  gb_ov:*
         6:  gb_pat:*
         7:  gb_ph:*
         8:  gb_pl:*
         9:  gb_pr:*
        10:  gb_ro:*
        11:  gb_sts:*
        12:  gb_sy:*
        13:  gb_uni:*
        14:  gb_vi:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

                          SUMMARIES
```

Result No.		Score	Query Match %	Query Length	DB	ID	Description
1	c	3291	100.0	2131	3	AY360470	AY360470 Caenorhab
2	c	2938	89.3	45389	3	U80032	U80032 Caenorhabdi
3	c	1873.5	56.9	17418	3	CBRG03O12	AC084485 Caenorhab
4	c	585	17.8	39089	3	CBRG33D04	AC084550 Caenorhab

No.		Score	Pred			ID	Description
5		181	5.5	3631	6	CQ575340	CQ575340 Sequence
6		181	5.5	3745	5	AF367177	AF367177 Drosophil
7		180	5.5	5894	3	ACMHC	Y00624 Acanthamoeb
8	c	179.5	5.5	19514	8	SPAC29E6	Z66525 S.pombe chr
9	c	179.5	5.5	36315	9	SPAC30	AL136538 S.pombe c
10		177	5.4	3543	6	AX151655	AX151655 Sequence
11	c	177	5.4	292967	14	AF369029	AF369029 White spo
12	c	177	5.4	305107	6	AX151396	AX151396 Sequence
13	c	177	5.4	305107	14	AF332093	AF332093 White spo
14		176.5	5.4	7396	6	AX686226	AX686226 Sequence
15		175.5	5.3	5122	9	HUMMYONM	M31013 Human nonmu
16		174	5.3	6788	5	XELNMUSMYO	L09740 Xenopus lae
17		173.5	5.3	6974	2	AF055895	AF055895 Xenopus l
18		173.5	5.3	7450	9	AB191263	AB191263 Homo sapi
19	c	173	5.3	307287	14	AF440570	AF440570 Shrimp wh
20		172	5.2	5670	3	TCA306290	AJ306290 Toxocara
21	c	171.5	5.2	110000	8	AE016615_4	Continuation (5 of
22		171	5.2	96210	1	YPPMT1	AL117211 Yersinia
23	c	171	5.2	100984	1	AF053947	AF053947 Yersinia
24		171	5.2	100990	1	AF074611	AF074611 Yersinia
25		171	5.2	100990	6	AR487637	AR487637 Amoeba pr
26	c	171	5.2	106642	1	AE017045	AE017045 Yersinia
27		170.5	5.2	4281	9	HSMYOIHMA	Z22957 H.sapiens m
28		170.5	5.2	5556	9	AK131080	AK131080 Homo sapi
29		169.5	5.2	2896	10	BC044834	BC044834 Mus muscu
30		169.5	5.2	5254	1	AK131171	AK131171 Mus muscu
31		169.5	5.2	5946	6	CQ730200	CQ730200 Sequence
32		169.5	5.2	6137	9	AB002376	AB002376 Human mRN
33		169.5	5.2	68727	3	AC004516	AC004516 Drosophil
34		169	5.1	6499	3	AF136711	AF136711 Drosophil
35		168.5	5.1	3858	6	CQ735115	CQ735115 Sequence
36		168.5	5.1	7355	10	MMU312390	AJ312390 Mus muscu
37		168.5	5.1	53424	3	AB055861	AB055861 Procambar
38	c	168.5	5.1	184704	2	CR385063	CR385063 Danio rer
39		168	5.1	1868	5	BC083371	BC083371 Danio rer
40		168	5.1	5439	6	CQ573099	CQ573099 Sequence
41		168	5.1	5756	6	CQ575339	CQ575339 Sequence
42		168	5.1	30386	2	AC014301	AC014301 Drosophil
43	c	168	5.1	134053	8	BX908812	BX908812 Neuropor
44	c	168	5.1	185405	3	AC023743	AC023743 Drosophil
45	c	168	5.1	299686	3	AE003444	AE003444 Drosophil

```
                          ALIGNMENTS

RESULT 1
AY360470
LOCUS       AY360470    2131 bp    mRNA    linear   INV 01-JUN-2004
DEFINITION  Caenorhabditis elegans osmotic stress resistance protein (osr-1)
            mRNA, osr-1-rml allele, complete cds.
ACCESSION   AY360470
VERSION     AY360470.1  GI:38154581
KEYWORDS    .
SOURCE      Caenorhabditis elegans
  ORGANISM  Caenorhabditis elegans
            Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
            Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE   1  (bases 1 to 2131)
  AUTHORS   Solomon,A., Bandhakavi,S., Jabbar,S., Shah,R., Beitel,G.J. and
            Morimoto,R.I.
  TITLE     Caenorhabditis elegans OSR-1 Regulates Behavioral and Physiological
            Responses to Hyperosmotic Environments
  JOURNAL   Genetics 167 (1), 161-170 (2004)
  PUBMED    15166144
REFERENCE   2  (bases 1 to 2131)
  AUTHORS   Solomon,A., Bandhakavi,S., Jabbar,S., Shah,R., Beitel,G.J. and
            Morimoto,R.I.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-AUG-2003) Biochemistry, Northwestern University, 2205
            Tech Drive, Hogan 2-100, Evanston, IL 60201, USA
FEATURES             Location/Qualifiers
     source          1..2131
                     /organism="Caenorhabditis elegans"
```

```
          /mol_type="mRNA"
          /db_xref="taxon:6239"
          /chromosome="I"
          /map="-0.2 m.u."
gene      1. .2131
          /gene="osr-1"
          /allele="osr-1-rm1"
5'UTR     1. .98
          /gene="osr-1"
          /allele="osr-1-rm1"
CDS       99. .2030
          /gene="osr-1"
          /allele="osr-1-rm1"
          /function="mediates multiple responses to osmotic stress"
          /note="OSR-1"
          /codon_start=1
          /product="osmotic stress resistance protein"
          /protein_id="AAR12229.1"
          /db_xref="GI:3815458 2"
          /translation="MILFLFLFLLLGFCIAPLSAQSPSTSDAPGALLSSLVGKSHQKL
          PLAPSMEALELMGVQFVDALIKKGQMEMAKGAPKTQLEVLEKVHPDQFDKYKKLKVDD
          LAADAVMQQAEMAKLQPKSGNAFIDMLNGNGIPIGSSIRGLEDAIRTQRDMENTDPSE
          QIAKAVMDKFQTQILPGLVANMIAGKNPFKMPQQMRKAQAPSSVFQQALAQRAMLGK
          NAPVAGRGEEQRMMNNRVDQRMQQRELQEEDEDDDLEDEDVPRRRSSDGEPQSEAE
          HQRRDLARRLKSSPRLKELLQNAEVQSLLSYQRMRDSPLSKRPLAMNDEDESAFRAM
          EARAKLDQKSQLVLGLHGPGESDDDEDBEDENLIDPSENSFRRAPLRLSSGFVEKLKS
          NDELKSALDRIKYRVDDVEKYLAPKPMEFNPKPQPGYFAPRKIPTRPRKMLPLLIGSD
          PKVQEEIRRHPSTEWKIAKESRVLTNLKNNPSLAALFMDDKLENTLKGRQMLTDEQKG
          RTRVKTIRALPRLFGAPTAKAEMIDAKVFQDIERPIPLFFEPKGRHTRLRWTGANE
          KEIPGLGSRFIPLPSLDPTMPALNTAFSTQGRARDEWDTMFKIPNNWNPGDEVGFKMNS
          KTKRFVGGNGAFDMPALGL"
3'UTR     2031. .2131
          /gene="osr-1"
          /allele="osr-1-rm1"
ORIGIN

Alignment Scores:
Pred. No.:            2131          Length:          2131
Score:            3291.00          Matches:          643
Percent Similarity:  100.00%       Conservative:       0
Best Local Similarity: 100.00%     Mismatches:         0
Query Match:         100.00%       Indels:             0
DB:                                Gaps:               0
```

US-10-736-868-2 (1-643) x AY360470 (1-2131)

```
Qy    1 MetIleLeuPheLeuPheLeuLeuLeuGlyPheCysIleAlaProLeuSerAla     20
Db   99 ATGATTTTATTTTATTTTTTATTCCTTGTTGTTGTTGTTTGTATCGCACCATTATCGGCC  158
Qy   21 GlnSerProSerThrSerAspAlaProGlyAlaLeuLeuSerSerLeuValGlyLysSer  40
Db  159 CAATCTCCATCGACTTCCGATGCTCCGGGAGCTTTGTTGTCATCTCGTAGGTAAAAGC    218
Qy   41 HisGlnLysLeuLysGluLeuLeuGlnAsnAlaGluValGlnSerLeuLeuSerTyr     60
Db  219 CATCAAAAACTACCACTGGCTCCATCAATGAAGCTCTTGAACTGATGGGTGTTCAATTT   278
Qy   61 ValAspAlaLeuIleLysLysGlyGlnMetGluMetAlaLysGlyAlaProLysThr     80
Db  279 GTTGATGCTCTCATCAAAAAGGTCAAATGAAATGGCAAAAGGAGCATTTAAGACTCAA   338
Qy   81 LeuGluValLeuGluLysValHisProAspGlnPheAspLysTyrLysLysLeuLysVal 100
Db  339 TTAGAAGTTCTAGAAGAAAGTACATCCTGATCAATTCGATAAGTACAAAAAGCTAAAAGTT 398
Qy  101 AspAspLeuAlaAlaAspAlaValMetGlnGlnAlaGluMetAlaLysLeuGlnProLys 120
Db  399 GATGATTTGGCAGCTGATGCAGTTATGCAACAGGCCGAGATGGCAAAATTACAGCCTAAA  458
Qy  121 SerGlyAsnAlaPheIleAspMetLeuAsnGlyAsnGlyIleProIleGlySerSerIle 140
Db  459 TCAGGAAATGCATTTATCGATATGTTGAATGGAAATGGAATCCCAATTGGTAGCAGTATT  518
```

```
Qy  141 ArgGlyLeuGluAspAlaIleArgThrGlnArgAspMetGluAsnThrAspProSerGlu 160
Db  519 CGTGGATTAGAAGATGCTATCCGAACGCAGCGAGATATGGAAAATACGGATCCGTCCGAA  578
Qy  161 GlnIleAlaLysAlaValMetAspLysPheGlnThrGlnIleLeuProGlyLeuValAla 180
Db  579 CAGATTGCCAAAGCCGTAATGGACAAATTCAAACAACAAATTCTCCCAGGACTCGTTGCA  638
Qy  181 AsnMetIleAlaGlyLysAsnProPheLysMetProGlnGlnMetArgLysAlaGlnAla 200
Db  639 AATATGATCGCTGGCAAGAACCCCTTTAAAATGCCTCAACAAATGAGAAAAGCTCAAGCT  698
Qy  201 AlaProSerSerValPheGlnGlnAlaLeuAlaGlnArgAlaMetLeuGlyLysAsnAla 220
Db  699 GCTCCATCGTCAGTGTTCCAACAAGCTCTTGCTCAAAGAGCAATGTTAGGTAAAAATGCC  758
Qy  221 ProValAlaGlyArgGlyGluGluGlnArgMetMetAsnAsnArgValAspGlnArg    240
Db  759 CCAGTTGCCGGTGGAAGAGGTGAAGAACAACGATGATGAATGGAGTGGAACAAAGA    818
Qy  241 MetGlnGlnArgGluLeuGlnGluGluAspGluAspAspAspLeuGluAspGlu      260
Db  819 ATGCAACAAAGAGAACTTCAAGAGGAAGATGAAGATGATGATCTTGAGGACGAGGAT    878
Qy  261 ValProArgArgArgSerSerAspGlyGluProGlnSerGluAlaGluHisGlnArgArg 280
Db  879 GTACCCAGAAGAAGAGAGTTCGGATGGAGAACCACAAAGTGAAGCAGCATCAGAAGAA   938
Qy  281 AspLeuAlaArgArgLeuLysSerSerProArgLeuLysGluLeuLeuGlnAsnAlaGlu 300
Db  939 GATTTAGCCAGGAGATTGAAAAGTCCTAGATTAAAAGAGCTTTTACAGAATGCGAA     998
Qy  301 ValGlnSerLeuLeuSerTyrGlnArgMetArgAspSerProLeuSerLysArgPro   320
Db  999 GTTCAATCTCTCTCTCTCTACCAACGAATGAGGATTCCACTGAGCAAGCGAAGGCCT   1058
Qy  321 LeuAlaMetAsnAspGluAspGluSerAlaPheArgAlaMetGluAlaArgAlaLysLeu 340
Db 1059 TTGGCTATGAACGATGAGGATGAAAGTGCATTCCGCGCAATGGAGGCTCGTGCAAAACTA 1118
Qy  341 AspGlnLysSerGlnLeuValLeuGlyLeuHisGlyProGlyGluSerAspAspAspGlu 360
Db 1119 GATCAAAAATCTCAACTTGTGCTCGGTCTCCATGGTTTGGAGAGTCTGATGATGATGAA  1178
Qy  361 AspGluGluAspGluAsnLeuIleAspProSerGluAsnSerPheArgArgAlaProLeu 380
Db 1179 GACGAAGAAGATGAAAATTTGATTGATCCATCTGAAAATTCATTCCGTGCGTCCACTT   1238
Qy  381 ArgLeuSerSerGlyPheValGluLysLeuLysSerAsnAspGluLeuLysSerAlaLeu 400
Db 1239 CGTCTTCTTTCCGGATTCGTTGAGAAATTAAAGTCAAATGAATTGAAAAGTGCATTG   1298
Qy  401 AspArgIleLysTyrArgValAspAspValGluLysTyrLeuAlaProLysProMetGlu 420
Db 1299 GACAGAATTAAATATCGAGTTGATGACGTGGAAAAGTATCTTGCTCCAAAGCCGATGGAA  1358
Qy  421 PheAsnProLysProGlnProGlyTyrPheAlaProArgLysIleProThrArgProArg 440
Db 1359 TTCAATCCAAAACCTCAGCCTGGCTACTTTGCTCCACGAAAATCCCACACAAGACCACGT  1418
Qy  441 LysMetLeuProLeuLeuIleGlySerAspProLysValGlnGluGluIleArgArgHis 460
Db 1419 AAAATGTTTCCATTATTAATTGGATCTGATCCAAAAGTTCAAGGAGAAACGAAGACAT    1478
Qy  461 ProSerThrGluTrpLysIleAlaLysGluSerArgValLeuThrAsnLeuLysAsnAsn 480
Db 1479 CCAAGTACCGAATGGAAAATTGCAAAGAATCAAGAGTTTTGACAAATTTGAAGAATAAT   1538
Qy  481 ProSerLeuAlaAlaLeuPheMetAspAspLysLeuGluAsnThrLeuLysGlyArgGln 500
Db 1539 CCAAGTCTTGCTGCATTGTTCATGGATGATAAACATTGAGAAATTAGAGAAAGGAAGGCCA 1598
```

```
Qy   501  MetLeuThrAspGluGlnLysGlyArgThrArgValLysThrIleArgAlaLeuProArg  520
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1599  ATGTTAACTGATGAACAGAAAGGTAGAACACGTGTCAAAACAATTCGTGCATTACCAAGA 1658

Qy   521  LeuPheGlyAlaProThrAlaLysAlaGluMetIleAspAlaLysValPheGlnAspIle  540
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1659  CTGTTCGGTGCGCACCAACTGCAAAAGCTGAAATGATTGATGCAAAGGTATTCCAAGATATT 1718

Qy   541  GluGluArgProIleProProLeuPhePheGluProLysGlyArgHisThrArgLeuArg  560
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1719  GAAGAAGTCCCATTCCTCCATTGTTCTTTGAACCAAAGGAAGGCATACGAGATTGAGA  1778

Qy   561  TrpThrGlyAlaAsnGluLysGlyIleLeuProGlyLeuGlySerArgPheLeuProSer  580
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1779  TGGACTGGAGCAAATGAAAGAAATTCCAGGACTTGGAAGTCGCTTCATTCTCCCATCT  1838

Qy   581  LeuAspProThrMetProAlaLeuAsnThrAlaPheSerThrGlnGlyAlaArgAlaArgAsp  600
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1839  CTTGATCCAACTATGCCAGCCTTGAACACGGCCTTTCTCGACTCAGGGCCGAGCCCGTGAC 1898

Qy   601  GluTrpProAspThrMetPheLysIleProAsnAsnTrpAsnProGlyValValGlyPhe  620
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1899  GAGTGGGATACCATGTTCAAAATCCCGAATAACTGGAATCCTGGAGATGAAGTTGGGTTC 1958

Qy   621  LysMetAsnSerLysThrLysArgPheValGlyValGlyAsnGlyAlaPheAspMetProAla  640
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1959  AAAATGAACTCAAAAACCAAACGATTCGTTGGAGAAATGGAGCATTGATATGCCTGCA  2018

Qy   641  LeuGlyLeu  643
          ||||||||||
Db  2019  CTGGGGATTG 2027
```

```
RESULT 2
U80032/c
LOCUS       U80032      45389 bp    DNA     linear   INV 16-SEP-2004
DEFINITION  Caenorhabditis elegans cosmid C32E12, complete sequence.
ACCESSION   U80032
VERSION     U80032.1  GI:2073526
KEYWORDS    HTG.
SOURCE      Caenorhabditis elegans
  ORGANISM  Caenorhabditis elegans
            Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
            Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE   1  (bases 1 to 45389)
  AUTHORS
  CONSRTM   WormBase Consortium
  TITLE     Genome sequence of the nematode C. elegans: a platform for
            investigating biology. The C. elegans Sequencing Consortium
  JOURNAL   Science 282 (5396), 2012-2018 (1998)
  MEDLINE   99069613
   PUBMED   9851916
REFERENCE   2  (bases 1 to 45389)
  AUTHORS   Wilcox,L.
  TITLE     The sequence of C. elegans cosmid C32E12
  JOURNAL   Unpublished (2001)
REFERENCE   3  (bases 1 to 45389)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-NOV-1996) Department of Genetics, Washington
            University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
            Louis, MO 63110, USA
REFERENCE   4  (bases 1 to 45389)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-MAY-1997) Department of Genetics, Washington
            University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
            Louis, MO 63110, USA
REFERENCE   5  (bases 1 to 45389)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-OCT-2001) Department of Genetics, Washington
            University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
            Louis, MO 63110, USA
REFERENCE   6  (bases 1 to 45389)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-APR-2002) Department of Genetics, Washington
            University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
            Louis, MO 63110, USA
REFERENCE   7  (bases 1 to 45389)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-JUL-2002) Department of Genetics, Washington
            University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
            Louis, MO 63110, USA
REFERENCE   8  (bases 1 to 45389)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-NOV-2002) Department of Genetics, Washington
            University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
            Louis, MO 63110, USA
REFERENCE   9  (bases 1 to 45389)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-JAN-2003) Department of Genetics, Washington
            University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
            Louis, MO 63110, USA
REFERENCE   10 (bases 1 to 45389)
  AUTHORS
  CONSRTM   WormBase Consortium
  TITLE     Direct Submission
  JOURNAL   Submitted (16-SEP-2004) Department of Genetics, Washington
            University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
            Louis, MO 63110, USA
            On May 8, 1997 this sequence version replaced gi:1703546.
COMMENT     Submitted by:
            Genome Sequencing Center
            Department of Genetics, Washington University
            St. Louis, MO 63110, USA, and
            Sanger Centre, Hinxton Hall
            Cambridge CB10 1RQ, England
            email: submissions@watson.wustl.edu and jes@sanger.ac.uk

            NOTICE:  This sequence may not be the entire insert of this clone.
            It may be shorter because we only sequence overlapping sections
            once, or longer because we provide a small overlap between
            neighboring submissions.

            This sequence was finished as follows unless otherwise noted: all
            regions were double stranded, sequenced with an alternate chemistry
            or covered by high quality data (i.e., phred quality >= 30); an
            attempt was made to resolve all sequencing problems, such as
            compressions and repeats; all regions were covered by sequence from
            more than one m13 subclone.

            For a graphical representation of this clone sequence and its
            analysis see:
            http://www.wormbase.org/db/seq/sequence?name=C32E12;class=Sequence

            NEIGHBORING CLONE INFORMATION

            The 5' clone is F56A3, 200 bp overlap; the 3' clone is B0261, 3201
            bp overlap. Actual start of this clone is at base position 1 of
            C32E12; actual end is at 42388 of C32E12.

            NOTES:

            Coding sequences below are the result of integration and manual
            review of the following data.: computer analysis using the program
            Genefinder (P. Green and L. Hillier, personal communication), the
            large scale EST projects of Yuji Kohara
            (http://www.ddbj.nig.ac.jp/c-elegans/html/CE_INDEX.html) and The C.
            elegans ORFeome cloning project (http://worfdb.dfci.harvard.edu/),
            similarity to other proteins from BlastX analyses
            (http://blast.wustl.edu/), sequence conservation with C. briggsae
            using Jim Kent's WABA alignment program (Genome Research
```

```
FEATURES     Location/Qualifiers
source       1. .45389
             /organism="Caenorhabditis elegans"
             /mol_type="genomic DNA"
             /strain="Bristol N2"
             /db_xref="taxon:6239"
             /chromosome="I"
             /clone="C32E12"

misc_RNA     750. .772
             /gene="mir-88"
             /product="micro RNA"
             /note="standard_name=C32E12.6"
             /db_xref="WormBase:C32E12.6"

gene         972. .2572
             /gene="C32E12.2"

CDS          join(972. .1126,1184. .1438,1501. .1587,1633. .1786,
             1997. .2325,2371. .2572)
             /gene="C32E12.2"
             /locus_tag="C32E12.2"
             /standard_name="C32E12.2"
             /note="contains similarity to Pfam domains PF01590 (GAF
             domain), PF00233 (3'5'-cyclic nucleotide
             phosphodiesterase); coded for by the following C. elegans
             cDNAs: OSTF082F3_1, OSTR082F3_1"
             /codon_start=1
             /product="Hypothetical protein C32E12.2"
             /protein_id="AAB53879.1"
             /db_xref="GI:1703547"
             /db_xref="WormBase:C32E12.2"
             /translation="MSKGIAGYVASTGEGLNIENAYEDERFNADVDSKTGYTTKTILC
             MPILIRGIVIGVVQMVNKHDGVFTRQDEDAFEIFAVCGLALHHAKLYDKIRRSEQKY
             RVALEVLAYHSVCNADEVNKLKLKKIEINNRIVELETIDPNGMRLSELEKPLYAVYMFKT
             LFADTLRPDTEDLIRFVLTVRKNYRRVAXHNWAHGWSVAHAMEFATLMNSPDAFTKLEA
             LALYVSCLCHDLDHRGKNNAYMKTMSTPLASIYSTSVMERHHFNQTVTILQQDGHNIL
             KSLSSEDYKKTLSLIKHCILATDLALFFSNKAKLNVILDNNTFDINRQEHRLLTQAVM
             MTGCDLVASAKPWNIQTETVKVIFEEFYDQGDAERLSGKEPIPMMDRRQQAHMLPQMQA
             S"

gene         complement(3240. .6294)
             /gene="osr-1"

CDS          complement(join(3240. .3454,3601. .3712,3759. .4390,
             4446. .4527,5034. .5441,5571. .5689,5742. .6059,6249. .6294))
             /gene="osr-1"
             /locus_tag="C32E12.3"
             /standard_name="C32E12.3"
             /note="contains similarity to Schizosaccharomyces pombe
             Hypothetical protein C29E6.10c in chromosome I.;
             SW:YAFA_SCHPO; coded for by the following C. elegans
             cDNAs: yk563c9.3, AY360470"
             /codon_start=1
             /product="Osmotic stress resistant protein 1"
             /protein_id="AAM15555.1"
             /db_xref="GI:20198782"
             /db_xref="WormBase:C32E12.3"
             /translation="MILFLFLFLLLGFCIAPLSAQSPSTSDAPGALLSSLVGKSHQKL
             PLAPSMEALELMGVGVFVDALIKKGQMEWAKGAFKTQLEVLEKVHPDQFDKYKKLVDD
             LAADAVMQQAEMAKLQPKSGNAFIDMLNGNGIPIGSSIRGLEDAIRTQRDMENTDPSE
             QIAKAVMDKFQTQILPGLVANMIAGKNPFKMPQQMRKAQAAPSSVFQQALAQRAMLGK
             NAPVAGGRGEEQRMMNNRVDQRMQQRELQEEDEDDDLEDDVPRRRSSDGPASNEDSAFRAM
             EARAKLDQKSQLVLGLHGFGESDDDEDEDENLIDPSENSFRRAPLRLSSGFVEKLKS
             NDELKSALDRIKYRVDDVEKYLAPKPMEFNPKPQPGYFAPRKIPTRPRKMLPLLIGSD
             PKVQEIRRHPSTEWKIAKESRVLTNLKNNPSLAALFMDDKLENTLKGRQMLTDEQKG
             RTRVKTIRALPRLFGAPTAKAEMIDAKVFQDIEERPIPPLFFEPKGRHTRLRWTGANE
             KEIPGLGSRFILPSLDPTMPALNTAFSTQGRARDEWDTMFKIPNNWNPGDEVGFKMNS
             KTKRFVGGNGAFDMPALGL"

gene         complement(8197. .11009)
             /gene="sox-1"
```

```
             /locus_tag="C32E12.5"
             complement(join(8197. .8424,9119. .9232,9281. .9688,
             10208. .10472,10809. .11009))
             /gene="sox-1"
CDS          /locus_tag="C32E12.5"
             /standard_name="C32E12.5"
             /note="contains similarity to Pfam domain PF00505 (HMG
             (high mobility group) box); coded for by the following C.
             elegans cDNAs: yk397g12.5, yk404e6.5, yk657g12.5,
             yk1577b07.5, yk1661e08.5, yk397g12.3, yk404e6.3,
             yk657g12.3, yk1577b07.3, yk1661e08.3"
             /codon_start=1
             /product="Sox (mammalian sry box) family protein 1"
             /protein_id="AAM1556.1"
             /db_xref="GI:20198784"
```

```
Alignment Scores:
Pred. No.:                              Length:              45389
Score:                   8.07e-181      Matches:             641
Percent Similarity:      2938.00        Conservative:        0
Best Local Similarity:   62.84%         Mismatches:          2
Query Match:             62.84%         Indels:              379
                         89.27%         DB:                  7
                         3

DB:

US-10-736-868-2 (1-643) x U80032 (1-45389)
```

```
Qy      1 MetIleLeuPheLeuPheLeuLeuLeuGlyPheCysIle------------  15

Db   6294 ATGATTTTATTTTATTTTTATTCCTGTTGTTAGGATTTTGTATCGGTTAGTTTTAAATT  6235

Qy     15 ------------------------------------------------------  15

Db   6234 CAATATTTAGAGATTACATTGTAAATTATGTTCAAAAAAAATCAGAAACTTTAAAAA  6175

Qy     15 ------------------------------------------------------  15

Db   6174 ATTTGGCAATTTATCAAATACTTGGCTGAACTTTTGTGATATCCTAACATTTTTTAA  6115

Qy     16 --------------------------------------------AlaPro  17

Db   6114 GTGATGATTTATGAGAGAAAATTGAAAAAGGGAAAATATTTAGAAAATTTCCAGCACCA  6055

Qy     18 LeuSerAlaGlnSerProSerThrSerAspAlaProGlyAlaLeuLeuSerSerLeuVal  37

Db   6054 TTATCGGCCCCAATCCCATCCGATGCTCCGGGAGCTTTGTGTCATCTCTCGTA  5995

Qy     38 GlyLysSerHisGlnLysLeuProLeuAlaProSerMetGluAlaLeuGluLeuMetGly  57

Db   5994 GGTAAAGGCCATCAAAAACTACCACTGGCTCCATCAATGGAAGGCTCTTGAACTGATGGGT  5935

Qy     58 ValGlnPheValAspAlaLeuIleLysLysGlyGlnMetGluMetAlaLysGlyAlaPhe  77

Db   5934 GTTCAATTTGTTGATGCTCTCATCAAAAAAGGTCAAATGGAAATGGCAAAAGGAGCATTT  5875

Qy     78 LysThrGlnLeuGluValGluLysValHisProAspGlnPheAspLysTyrLysLys  97

Db   5874 AAGACTCAATTAGAAGTTGAGAAAGTACATCCTGATCAATTCAATTACAAAAAG  5815

Qy     98 LeuLysValAspAspLeuAlaAlaAspAlaValMetGlnGlnAlaGluMetAlaLysLeu  117

Db   5814 CTAAAAGTTGATGATTTGGCAGCTGACGCTGTATGCAACAGGCGGAGATGGCAAAATTA  5755

Qy    118 GlnProLys------------------------------------------------  120

Db   5754 CAGCCTAAATCAGGTTTCATTTTACTTTTCAGTAACTGTTCTGAAAACAAATTTGAAT  5695

Qy    121 -SerGlyAsnAlaPheIleAspMetLeuAsnGlyAsnGlyIleProIleGlySerSerI  140

Db   5694 TTCAGGAAATGCCATTATCGATATGTTGAATGGAAATGGAATCCCAATTGGTAGCAGTAT  5635

Qy    140 eArgGlyLeuGluAspAlaIleArgThrGlnArgAspMetGluAsnThrAspProSerGl  160

Db   5634 TCGTGGATTAGAAGATGCTATCCGAACACAGCGAGAGATATGGAAAATACGGATCCGTCCGA  5575
```

```
Qy    160 uGln---                                                          161
              |||
Db   5574 ACAGGTTGGTGTGAATTGCAGTTTAAAAAATATTAGAGAAACTCCCTTTTCCCA         5515

Qy    161 ------------------------------------------------------------     161
Db   5514 TTTGTTTGGGTTTTAATTATTGTTGCAGAATGATTTTCTTAATTTTATAACTAAACAAT    5455

Qy    162 ------------IleAlaLysAlaValMetAspLysPheGlnThrGlnIleIleLeuProGl  177
Db   5454 GTTTCTAATTTAGATTGCCAAAGCCGTAATGGACAAAATTCAAACACAAATTCTCCCCAGG  5395

Qy    177 yLeuValAlaAlaMetIleAlaGlyLyLysAsnProPheLysMetProGlnMetArgGlY    197
Db   5394 ACTCGTTGCAAATATGATCCGCTGCAAGAACCGTGATGATGAGACGAAGAGCAATGTTAGG  5335

Qy    197 sAlaGlnIleAlaAlaProSerValPheGlnGlnArgLeuAlaLeuAlaArgAlaMetLeuG  217
Db   5334 AGCTCAAGCTGCTCCATCGTCAGTGTTCCAACAAGCCTTGCTCAAGAGCAATGTTAGG     5275

Qy    217 YLysAsnAsnAspGluLeuLeuGlyArgGlyGluGluGlnAsnArgMetMetAsnArgVa   237
Db   5274 TAAAAATGCCCCAGTTGCCCGGTGGAAGAGGTGAAGAACAACCGGATGATGAATCGAGT    5215

Qy    237 lAspGlnArgMetGlnIleArgGlnLeuGlnLeuGlnGluGlnAspAspAspAspLeuGl   257
Db   5214 GGACCAAAGAATGCAACAAAGAGAACTTCAAGAGGAAGATGAAGGATGATGATCTTGA     5155

Qy    257 uAspGluAspValProArgArgArgSerSerAspGlyGluProGlnSerGluAlaGluHi   277
Db   5154 GGACGAGGATGTACCCAGAGAAGAGAAGTTCGGATGGAGAACCACAAAGTGAGCAGAGCA   5095

Qy    277 sGlnAsnArgArgAspLeuAlaArgLeuLysSerSerProArgLeuLysGluLeuLeu--   296
Db   5094 TCAGAGAGAGATTTAGCCAGGAGATTGAAAAGTAGTCCTAGATTAAAAGAGCTTTT-AC   5036

Qy    296 ------------------------------------------------------------   296
Db   5035 AGGTGAATTAGTAATTTCTCAACTGTTTTAAATTTTTTTAGAAAGGAAAAAAAATCCAAA   4976

Qy    296 ------------------------------------------------------------   296
Db   4975 ATATTTAAGCTTTATATAAAAGTATTAAGTCTTTTATATGTATTTAGAGCATATATCTGTA  4916

Qy    296 ------------------------------------------------------------   296
Db   4915 AATATAATTTTACAAATACAACATCATCACATAGGTGTTGAGAATCGCCGATATTGATAA   4856

Qy    296 ------------------------------------------------------------   296
Db   4855 CTTTCAACAGTCGGTAATACTTCTGCAGCAATTATCAAAAAAATTGCCGAATCTTACTAT   4796

Qy    296 ------------------------------------------------------------   296
Db   4795 TTTTAAGCTTCTATATTTTTCGGAAACTTCTAAAAAAGTTTTTTGTCATTTTGAAAATTTA  4736

Qy    296 ------------------------------------------------------------   296
Db   4735 ATATATAGGAATGCGATTTAGCTAAGTCTCCTATTTGTTTGTTGTAAGATGTTAAACCACC  4676

Qy    296 ------------------------------------------------------------   296
Db   4675 ATAAAATATTTCCAAAAATTTTGTGGAGATCCGGAATTGAGAAAAAAATCACGTGTATCT   4616

Qy    296 ------------------------------------------------------------   296
Db   4615 TGGAAAATAACTAACAATGACAGTCGTAAAGCCTAGTAGGACATTTGTGCCTGTGTGGTC   4556

Qy    297 ------------GlnAsnAlaGluValGlnSerLeuLeuSerTyrGl                308
Db   4555 GCTTTGAAAAAAATAACCTAATACTTTCAGAATGCCGAAGTTCAATCATTGCTTCTTACCA  4496
```

```
Qy    308 nArgMetArgAspSerProLeuSerLysArgArgProLeuAlaMetAsn------       324
Db   4495 ACGAATGAGGGAATTCTCCACTGAGCCAAGCGAAGGCCTTTGGCTATGAACGGTTGGTTTTT 4436

Qy    325 ------------------------------AspGluSerA                       330
Db   4435 GTTTTTGAATTAGTAGGAACTCTTATCTGTCGACTAAATTCAGATGAGGATGAAAGTG     4376

Qy    330 laPheArgAlaMetGluAlaArgAlaLysAlaAspGlnLysSerGlnLysValLeuGlyLyL  350
Db   4375 CATTCCGGCCAATGGAGGCTCGTGCAAAACTAGATCAAAAATCTCAACTTGTCTCGGTC    4316

Qy    350 euHisGlyPheGlyGluSerAspAspAspAspAspGluGluGluAsnGluIleuAspP     370
Db   4315 TCCATGGTTTTGGAGAGTCGTGATGATGAGACGAAGAGATGAATTGATTGATC         4256

Qy    370 roSerGluAsnSerPheArgArgAlaProLeuArgLeuSerGlyPheValGluLysL     390
Db   4255 CATCTGAAAATTCATTCCGTCGTGCCACCACTTCGTCTTTCCCGATTCGTTGAGAAAT    4196

Qy    390 euLysSerAsnAspGluLeuLeuSerAlaLeuAspArgIleLysTyrArgValAspV     410
Db   4195 TAAAGTCAAATGATGAAATGTGCATTGGACAGAATTAAATATCCAGTTGATGACG       4136

Qy    410 alGluLysTyrLeuAlaProLysProMetGluPheAsnProLysProGlnTyrP        430
Db   4135 TGAAAAGTATCTTGCTCCAAAGCCGATGGAATTCAATCCAAAACCTCAGCCTGCTACT    4076

Qy    430 heAlaProArgLysIleProThrArgProArgLeuLeuIleGlySerA              450
Db   4075 TTGCTCCACGTAAAATCCCAACAAGACCACGTAAAATGCTTCCATTATTAATTGGATCTG  4016

Qy    450 spProLysValGlnGluGluIleArgArgHisProSerThrGluTrpLysIleAlaLysG   470
Db   4015 ATCCAAAGTTCAAGAGGAAATACCGAAGACATCCAAGTACCGAATGGAAAATTGCAAAAG  3956

Qy    470 luSerArgValLeuThrAsnLeuLysAsnAsnProSerLeuAlaAlaLeuPheMetAspA   490
Db   3955 AATCAAGAGTTTTGACAAATTTGAAGAATAATCCAAGTCTTGCTGCATTGTTCATGGATG   3896

Qy    490 spLysLeuGluAsnThrLeuLysGlyArgGlnMetLeuThrAspGluGlnLysGlyArgT   510
Db   3895 ATAAATTAGAGAATACATTGAAAGGAAGGCAAAATGTTAACTGATGAACAGAAAGGTAGAA  3836

Qy    510 hrArgValLysThrIleArgAlaLeuProArgGluGlyAlaProThrAlaLysAlaG      530
Db   3835 CACGTGTCAAAACAATTCGTGCATTACCAAGACTGTTCGGTGCACCAACTGCAAAAGCTG  3776

Qy    530 luMetIleAspAlaLys---------------------------------------      535
Db   3775 AAATGATTGATGCAAAGTATAATGAGTATAATAATGAGATTAATTTATTTTGGCAATTAAATATATTATTT 3716

Qy    536 ---ValPheGlnAspIleGluProLeuPheGluProLysGly                    554
Db   3715 CAGTTATTCCAAGATATTGAAGAACGTCCCATTCCTCCATTTGAACCAAAAGGA        3656

Qy    555 ArgHisThrArgLeuArgTrpThrGlyValAlaAsnGluLysGluIleProGlyLeu----  572
Db   3655 AGGCATACCAGAGTTGAGATGGACTGGGAGCAAATGAAAAGAAATTCCAGGACT-TGGTAG  3597

Qy    572 ------------------------------------------------------------   572
Db   3596 GTATTGATTAAAAATGTTTTTGTCTTGGTTTTTCTGCAATGTCAGTGAAAACATAGAATGT  3537

Qy    572 ------------------------------------------------------------   572
Db   3536 GTACGAAGAACTATGTTACGTGCCGACAACAGAAATTGATGAGTCAAAAACAGAGAAAA    3477

Qy    573 ------------GlySerArgPheIleLeuProSerLeuAspProThrMet           585
Db   3476 AACGCCAATCCTTTTCTTATAGGAAGTCGCTTCATTCTCCCATCTTGTATCCAACTATG    3417

Qy    586 ProAlaLeuAsnThrAlaPheSerThrGlnGlyArgAlaArgAspGluTrpAspThrMet   605
```

```
Db  3416 CCAGCCTTGAACACGGCTTTCTGACTCAGGGGCGAGCCCCTGACCGAGTGGGATACCATG 3357
              ||||||||||||||||||||||||||||||||

Qy   606 PheLysIleProAsnTrpAsnProLysGlyGluValGlyPheLysMetAsnSerLys 625
Db  3356 TTCAAAATCCCGAATAACTGGAATCCTGGAGATGAAGTTGGGTTCAAAATGAACTCAAAA 3297
              ||||||||||||||||||||||||||||||||

Qy   626 ThrLysArgPheValGlyGlyAsnGlyAlaLeuGlyLys 643
Db  3296 ACCAAACGATTCGTTGGAGGAAATGGAGCATTTGATATGGCCACACTGGGATTG 3243
              ||||||||||||||||||||||||||||||||
```

RESULT 3
CBRG03012/c

```
LOCUS       CBRG03012              17418 bp    DNA     linear   INV 04-NOV-2000
DEFINITION  Caenorhabditis briggsae cosmid G03012, complete sequence.
ACCESSION   AC084485
VERSION     AC084485.1  GI:11094935
KEYWORDS    HTG.
SOURCE      Caenorhabditis briggsae
ORGANISM    Caenorhabditis briggsae
            Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
            Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE   1  (bases 1 to 17418)
AUTHORS     Washington University Genome Sequencing Center.
TITLE       The C. briggsae Genome Sequencing Project
JOURNAL     Unpublished
REFERENCE   2  (bases 1 to 17418)
AUTHORS     Wilson,R.
TITLE       The sequence of C. briggsae cosmid G03012
JOURNAL     Unpublished
REFERENCE   3  (bases 1 to 17418)
AUTHORS     Waterston,R.
TITLE       Direct Submission
JOURNAL     Submitted (04-NOV-2000) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT     Submitted by:
            Genome Sequencing Center
            Department of Genetics, Washington University,
            St. Louis, MO 63110, USA
            e-mail: jspieth@watson.wustl.edu

            NOTICE: This sequence may not be the entire insert of this clone.
            It may be shorter because we only sequence overlapping sections
            once, or longer because we provide a small overlap between
            neighboring submissions.
FEATURES             Location/Qualifiers
     source          1..17418
                     /organism="Caenorhabditis briggsae"
                     /mol_type="genomic DNA"
                     /strain="GujArat G16"
                     /db_xref="taxon:6238"
                     /clone="G03012"
ORIGIN
```

Alignment Scores:

Pred. No.:	7e-112	
Score:	1873.50	Length: 17418
Percent Similarity:	28.06%	Matches: 522
Best Local Similarity:	26.48%	Conservative: 31
Query Match:	56.93%	Mismatches: 24
DB:	3	Indels: 1396
		Gaps: 11

US-10-736-868-2 (1-643) x CBRG03012 (1-17418)

```
Qy    62 AspAlaLeuIleLysGlyGlnMetGluMetAlaLysGlyAlaPheLysThrGlnLeu 81
              ||||||||||||||||||||||||||||||||
Db 17417 GATGCTTTAATTAAAAAGACAGATGGAAATGGCAAAAGGAGCATTCAAAACCAACACTG 17358

Qy    82 GluValLeuGluLysProAspGlnPheAspLysTyrLysLeuLysValAsp 101
              |||||||||||||||||||||||:::
Db 17357 GAGGTTCTGGAGGAAAGTACATCCGGATCAATTTGAGAAATATAAGAAACTGAAGTGGAG 17298

Qy   102 AspLeuAlaAspAlaValMetGlnGlnAlaGluLysLeuGlnProLys--- 120
              ||||||||||||||||||||||:::||
Db 17297 GATTAGCGGCAGATGCAGTTATGCAACAGGCAGAAGTGTCAAGTTGCAACCGAAAACT 17238

Qy   120 120
Db 17237 GGTAAGGAGAATATAAAAGAGAGATCATTACTACGGAAGCAATACGACTGAAAATTATT 17178

Qy   120 120
Db 17177 TCTAGTTAGCACATGCAACAGTAAAAATTCAAAATTCAGTTTGAGATAATGTCAGAAA 17118

Qy   120 120
Db 17117 AAAGTTTTTATTTTCACATTTTAATAAATGTAAGTATCAATTTCCGATCGAATCAATTG 17058

Qy   120 120
Db 17057 GCAAATCATTAGAGGAATAAGTTTCTGGAAACTGAAAGTTTCAAATGTCAGATCTTG 16998

Qy   120 120
Db 16997 AAACTGAACTTTCGAAAGTCTGAGAAAACCTTAAGTTAAACATTGAACACTGTCTAATAG 16938

Qy   120 120
Db 16937 ACAGGGCTCCTAACTTCAAGTTTTCGAGACTTTATATTTTGGTGAAAAATGAGTAATA 16878

Qy   120 120
Db 16877 GTTCCATCCGGGAACTTAGTTCCGACTTTCAAGTTTAAAAATTCGTCTTCCCAAAACTT 16818

Qy   120 120
Db 16817 TAAATCGTACAAGACAGAAATTGTATGTATGAAGACCAAGAGTGCACTTACAATATATA 16758

Qy   121 SerGlyAsn 123
              ||||||||
Db 16757 AAAACCTAATAACCGCCGAAACAGTTATTTCTAGTCGACTATCTTTTCAGGAAAC 16698

Qy   124 AlaPheIleAspMetLeuAsnGlyAsnGlyIleLeuProIleGlySerSerIleArgGlyLeu 143
Db 16697 GCATTCATCGATATGTTAAATGGAAATGAAGAATTCCAATTGGAAGTAGTATTCTGTGGATTA 16638

Qy   144 GluAspAlaIleAlaIleArgThrGlnArgAspMetGluAsnThrAspProSerGlu--- 160
Db 16637 GAAGATGCTATTCGGACACCAAGAGACATGGAGAACACTGATCCATCTGAGCAGGTAATT 16578

Qy   160 160
Db 16577 TCACTTTTTAAGTTTGAATGACAGTCTATTGTGAAAATATGTAAATGTCATTCATGTTA 16518

Qy   160 160
Db 16517 TTCTTGAAACATAACATTTTTTCGTGCGAGTTCACCCGAGTTCATATGAAAGGTACCTA 16458

Qy   160 160
Db 16457 TGGCTCATTTGAAGGAGAAAGCGTTCTACAGTGCTATAAATAAAGCATAGTTACAAAAG 16398

Qy   160 160
Db 16397 GCCGGGCCGGGCCGGTACGAGAATTTTCTGGCCCGGCCCGCGGGCCCGCCTACCA 16338

Qy   160 160
Db 16337 CTGTTGAAAAATTTCCAATGAAAAATTGGATTTTTTTTGAATAAAAAGAAAAAAGGT 16278

Qy   160 160
Db 16277 TACATTTTAAACAAAATTAACAAAATAAAAAATAGTCAGAAATAGTCGTTTTGTATAT 16218

Qy   160 160
Db 16217 160
```

This page is a dense nucleotide/protein sequence alignment printout consisting of paired "Db" (database) and "Qy" (query) rows with position numbers, arranged in two columns. The alignment blocks include numbered coordinates such as:

Left column (Db/Qy):

```
Db  16217 ATTTTAATTCAAAAATTTCGAAAAAAAAATTCAAAAAAAT 16158
Qy    160 ---------------------------------------  160
Db  16157 CAGCCGTCTGTACAGACCCCGGGCCGTACTCCATGGGCCCGGCCCGTCAAAATTC 16098
Qy    160 ---------------------------------------  160
Db  16097 TCGACGGGCCGATCCGCGGCCGCCCGAGATTTGCAACTATGAAATAAAATGCTTCTA 16038
Qy    160 ---------------------------------------  160
Db  16037 CTTATGACGCCACATTTATCTGCACGCTATTGCACAAAGTTATTACCGGCATATTCAAA 15978
Qy    161 ---------------------------GlnIleAlaLysAlaValMe 167
Db  15977 AATTCACTGAAACTGCAATACAAATTTTAACATTCTTCCAGATTGCCAAAGCAGTCCT 15918
Qy    167 tAspLysPheGlnThrGlnIleLeuProGlyLeuValAlaAsnMetIleAlaGlyLysAs 187
Db  15917 TGACAAGTTTCAAACTCAAATTCTGCCTTGGACTCGTTGCCAATATGAATCGCCGGAAGGAA 15858
Qy    187 nProPheLysMetProGlnMetArgLysArgLysAlaGlnAlaAlaProSerValPheGl 207
Db  15857 TCCATTCAAAATGCCACAACAAATGAGAAAAGCCAAGCAGCCATCAGCAGTTCTTCA 15798
Qy    207 nGlnAlaLeuAlaGlnArgAlaMetLeuGlyLysAsnAlaProValAlaGlyValGlyArgGl 227
Db  15797 ACAAGCAATTGCTCAGCCGGAGAAATGCTTCCAGAAATG-----AGGAA 15744
Qy    227 yGluGluGlnRgMetMetMetAsnArgV21AspVa1GlnIleGlnArgGlu---Le 246
Db  15743 TAATCAACAAATGATGATGATGAATCAGTCGAAGATCGAATCGAACAGAGGAACAACT 15684
Qy    246 uGlnGluGluAspGluAspAspAspLeuGlnAspValProArgArgSe 266
Db  15683 AGTCGAAGAAGACGATGATGATGAAGTT---GAAGATGTACCCAAGAGAGAAGTTC 15630
Qy    266 rSerAspGlyGluProGlnSerGluAlaGluHisGlnArgArgAspLeuArgArgArgLe 286
Db  15629 TTCAAATGGAGAGCCACAAACTGAACGAGAGCACGAGAAAGGGATTTGATGAGGAGATT 15570
Qy    286 uLysSerProArgLeuLysGluLeu 296
Db  15569 GCCGGAATAGTCCAAGATTGAAGGGTCTTCTTCAGGTGAGTTATGGTTTTTAGACAAATAT 15510
Qy    296 ---------------------------------------  296
Db  15509 TTGACTATCGTAAAGCTGTAAAATACGATTTCGTGAAACATTTTTTAAAGGGGGGGGG 15450
Qy    296 ---------------------------------------  296
Db  15449 GGGACAGGAGATTTTTTGTAGTATTTGTGTAGCGTATGCATTCTACCGTGACGCGTG 15390
Qy    296 ---------------------------------------  296
Db  15389 TACAACGCTTAATGTCACTCTGCGTACAGCGGTTGTCCTCACCTTTATGAATGCAATCGG 15330
Qy    296 ---------------------------------------  296
Db  15329 ACGCCGGCTGCCTCTTGCCCGCTGCGTCAGTCGGTAGCGGCGCCGGCGAATACTAGTCTC 15270
Qy    296 ---------------------------------------  296
Db  15269 GACAAGTTCAACTCCTCGAGGGCATATATTAAACGAAATTTCGAGAACATTATTTTTTT 15210
Qy    296 ---------------------------------------  296
Db  15209 CATGTCCCCCACCTTTAAAGTTGACAGTCGGGATGCTTGGGCACAAAAACAGACTTCTG 15150
Qy    296 ---------------------------------------  296
```

Right column (Db/Qy):

```
Db  15149 TTTGAATCTTCTATTCCGAGCTAGCTAAATTTGTTTTCGTCGCTCTTTAGAGAACTATA 15090
Qy    297 --------------------------GlnAsnAlaGluValGlnSerLe 304
Db  15089 CTACATATTAACAAGGATTTCTAAATTTCATAGTTTCAGAACGCCGAGTGCAATCTCT 15030
Qy    304 uLeuSerTyrGlnArgMetArgAspSerProLeuSerLysArgProLeuAlaMetAs 324
Db  15029 TCTCTCTTACCGTCGAATGAAGAGAGACTCTCCACTCAGCAAACGAGGGCCATTGGCTATGAA 14970
Qy    324 ---------------------------------------  324
Db  14969 TGGTTCGTTCCTTTTTTTTCGAAAACGTTTTAAGGATTTTTATGAAGTTTTGGAGAAGTTA 14910
Qy    324 ---------------------------------------  324
Db  14909 TACTTCTGCATTTTTTAATTTTTTTTTCAAATTAGGACTTGCAGAGAGGATTTCGATTTTT 14850
Qy    324 ---------------------------------------  324
Db  14849 TATTGGGTATTAATCATTTAAAGGTGGGGGACATGAAAAAAAAATGTTCTCGAAATTCC 14790
Qy    324 ---------------------------------------  324
Db  14789 GTTCAAAGATGCCCCTCGAGGAGTTGAACTTGTGAAGGCCAGGATTGAGGGTACGGCGCC 14730
Qy    324 ---------------------------------------  324
Db  14729 GCTACCGACTGAGCCACGCGGGCAAAGAGAAAGGCGCACGCCCGATTGCATACATAAAGGC 14670
Qy    324 ---------------------------------------  324
Db  14669 GAGGACGACAGATCGTACGGAGAGTGTACACCGTGGTTGTACACCGTGACCGGTAGAATGCA 14610
Qy    324 ---------------------------------------  324
Db  14609 TACTTCTAGCTACAAAAAAACTACGAAAAAATTCCTGTCCCACCTTTAAGAAAAATCTT 14550
Qy    324 ---------------------------------------  324
Db  14549 GATTTGTTTCAATCTGAAAAATCGAAGTTTTTTTTTCAAGTACGTTCAAGCTTTCTCC 14490
Qy    324 ---------------------------------------  324
Db  14489 TATATTTCTTCTTCTTTGAAAATGACAGTCTGAACTTTGTATTATTAACTATTATCTAAATTTG 14430
Qy    324 ---------------------------------------  324
Db  14429 GATTTTCAAACTTTCTAAAAACGGTAAAAATTTTGAAAAAAACAAAAGTTACGACACA 14370
Qy    324 ---------------------------------------  324
Db  14369 AAAGAATCACCAAGTGCGATTTCTTCTTCTAAAAAATACGGTCATGATTTTTCCGTTTTTTTA 14310
Qy    324 ---------------------------------------  324
Db  14309 AAACTAATATTTTTGAATTAAAACAATTTCTGAAGTTCAATCTTTGCATTTTTCTAACAT 14250
Qy    324 ---------------------------------------  324
Db  14249 CCGTTTCTTCAACTTTTTAGAAACGGTGAAAAACTTTAAAAATGCTAACCACTGAGAGCA 14190
Qy    325 ------AspGluAspGluSerAlaPheArgAlaMetGluAlaArgAlaLysLeuAspGl 342
Db  14189 CTTTTTCAGATGAAGAACGAAAACATATATTCCGTGCGATGGAAGCCCGTGCAAAGCTCGATCA 14130
Qy    342 nLysSerGlnLeuValLeuGlyLeuHisGlyPhe------------GlyGluserAspAspAs 359
Db  14129 AAAATCGCAATTAGTTCTCGAATTCCTCCATGGATTCCATGGATTCCGGAATGGGAGATGAAGATGATGA 14070
Qy    359 pGluAspGluGluAspGluAsnLeuIleLeuAspProSerGluPheArgArgAlaPr 379
Db  14069 C---GATGAGGAAGATGAGAATTCAATTGATCCATCAGCCAACTCATTCCGTTCGTGCTCC 14013
```

Each alignment block pairs a "Db" reference sequence row with a "Qy" query sequence/translation row, with flanking coordinate numbers. The vertical bars between rows (not fully reproducible here) indicate base/residue identity.

```
Qy    379  oLeuArgLeuSerGlyPheValGluLysGluLeuLysSerAsnAspGluLeuLysSerAl  399
                ||||||||||||||   |||  |::|||||||
Db  14012  ACTTCGATTGCTTCCTCGGATTCGTTGAAAAACTAAAATCAAATGATGAATGAGAAGTGC  13953

Qy    399  aLeuAspArgIleLysTyrArgLeuAlaGluValGlyLysTyrLeuAlaProLysProMe  419
Db  13952  ATTGAACAGAATCAAGTACGAGTTGACGATGCTGGAAAAGTATTTGCAACCTAAGCCAAT  13893

Qy    419  tGluPheAsnProLysProGlnProGlyTyrPheAlaProArgLysIleProThrArgPr  439
                |::||||||||||
Db  13892  GGATTTCAATCCTAAACCCCAGCCTGGATATTTTGTACCTCGAAAAATCCCAACTAGACC  13833

Qy    439  oArgLysMetLeuProLeuLeuGlySerAspProLysValGlnGluGluIleGluArgAr  459
                ||||||||  |::|||
Db  13832  TCGTAAAATGCTTCCGTTGTTGATCGGATCAGATCCGAAAGTTCAAGAAGAGATCGAAAG  13773

Qy    459  gHisProSerThrGluTrpLysIleLeuAlaLysGluSerArgValLeuThrAsnLeuLysAs  479
Db  13772  ATATCCAAGCACAGAATGGAAGATTCTAGCAAAAGAAGTTCTGACAAATTTGAAGAA  13713

Qy    479  nAsnProSerLeuAlaAlaLeuPheMetAspAspLysLeuGluAsnThrLeuLysGlyTyr  499
                 |||||||::|||
Db  13712  CAATCCAAGTCTTGCTGCACTTTTTATGGATGAGAAAATACGTTAAGAGGAAG  13653

Qy    499  gGlnMetLeuThrAspGluGlnLnLysGlyArgPheArgValLysThrIleArgAlaLeuPr  519
                ||::|||
Db  13652  GCAATTGTTGACTGATGAGCAGAGAGGGAAGAACCCAGAGTTAAGACAATCAGAGCAATGCC  13593

Qy    519  oArgLeuPheGlyGlyAlaProThrAlaLysAlaGluMetIleAspAlaLysValValPhe--  537
Db  13592  GAGATTGTTTGGAGCACCAACTGCGAAAGCCGAGATGATCGATGCAAAGGTAAG-GAACA  13534

Qy    537  --------
Db  13533  GAAAAAAAGAGTGTTGTTTTCGGGAGTTTCTTAGGAACATGGAACATAGAGCAGTTGAATAG  13474

Qy    537  --------
Db  13473  AACATGCATTAGTTTTCAAATGTGAGTGTGTGAAACAGAAAAAAACGTAATTTCAAAATT  13414

Qy    537  --------
Db  13413  TGAAACGTGTTGCTTCTTCAATCATGCGCATCCAAAATACAAAAGACTGTAATTTCTTG  13354

Qy    537  --------
Db  13353  AGTAAAAAAATAGTCCGCCCCATATTTAACAACTTTCGTAAACCGACAGGAATATTTTCAA  13294

Qy    537  --------
Db  13293  TTTCACAATCAGTAGAATTTTCGGATTTTGTGTTCTTAATTTAAAATTTTTACAGCTGTTT  13234

Qy    537  --------
Db  13233  TATTTCGACTTTTTAGCAAAAATAGATTCTAATGATGTTTTAGAACTTTGAGAAATCTA  13174

Qy    537  --------
Db  13173  AAAATTCTAACAAAATAAAAATAATAACTGATATACAACACTATAGAAGCCCTCGATG  13114

Qy    537  --------
Db  13113  TTGTAATAGTTGCTGCTGTCATAAGCATCGTATCAACAATTGACATTAAATAAATTTCGCACT  13054

Qy    537  --------
Db  13053  TTTGGAGATATTCTGTTTCCGAAACAGGATAAAAAAAGGACCTAATTTTCCATTTTTCAGA  12994

Qy    537  --------
Db  12993  CATACTGAAAAATCGAAAACCTCGGTAAATATATAAATGTCCCTGTTCCAACTAATTTTT  12934
```

```
Qy    537  --------
Db  12933  TATTATTGCTTGACACTCATCGTTAATTGGTACAAAGCTGTTCTTCAACAATCTAATA  12874

Qy    537  --------
Db  12873  TCACAGGTTTCACCGTATCTTTCATAAACTTGGTATCAGCAACTAACACATTGAACGACCT  12814

Qy    537  --------
Db  12813  CCGACTATCTGCAATTAATTATTTCGAAACAGAATAAGTTGAAACATGCTCAAAACTAA  12754

Qy    537  --------
Db  12753  AAACCTATGTTCAAGTAAATTCCCTGTTCCACAGTTAATTTTGGTTATTGCTAGCTGGG  12694

Qy    537  --------
Db  12693  AGTCATGGTTGTTGTGTCTACATGCAAACGGAAAAATTCAGTATTTCAGGTCATTCTCTT  12634

Qy    537  --------
Db  12633  CAAACCTACAAGAAAAGTTTCATTAAATTTAATATTTAAAAAAACGAGATTTGGAGGTT  12574

Qy    537  --------
Db  12573  CAGAAAAGTAGTGTAATTCCAATAGGGCATACAGGATGTCGATGTTCATTGTTCCGTC  12514

Qy    537  --------
Db  12513  TGAAAAGAAACCTTTAGAAGGAAATTTATTTAAAATCGCTTCAAAAGCCTAGATGTTG  12454

Qy    537  --------
Db  12453  TAAGCTTTCATAAACTGCGGTATCCACCTACTTGAAATCGAATAAAATTGAGCAATTAAAAAA  12394

Qy    537  --------
Db  12393  AAGTAAGGGCCCGAAGTTGCATGTGAGCACACAACCATGACTCCCAGCTAGCAATAACC  12334

Qy    537  --------
Db  12333  AAAAATTAACTGTGAAACAGGGGAATTTACTTGAAAACTGTGAAAACACAGGGAATTTACTTCA  12274

Qy    537  --------
Db  12273  AACATCATGAAAATTGAAAAACATTGCTTCAATAGTTCAATAGTTACCCCGGTTTCACATATAATTTTTGG  12214

Qy    537  --------
Db  12213  TTATTGCTTGAAAACTTATCGTTAGTTCCGAATTGAAAACAAAAAATGCCAACTATTTCA  12154

Qy    538  ------GlnAspIleGluArgProIleProProLeuPhePheGluProLysGlyAr  555
                   |||||||||||||::::
Db  12153  GGTTACTTCAAGATATCGAAGACAACGACCGGTACCACCATTGTTCTTCGAGCCAAAAGGAA  12094

Qy    555  gHisThrLeuArgTrpThrGlyThrGlyValAlaAsnGluLysGluIleProGlyLeu-----  572
                :|||||
Db  12093  ACACACTCGGTTGAGATGGACTGGAGCCAATGGAGAAAGAAATTCCGGGACT-CGGTGGGT  12035

Qy    572  --------
Db  12034  TTCTTTTAAAAAATTGAATATGTAGAAAGTAGCTGAGAAGAACTATATTGAAAAACCAA  11975

Qy    572  --------
Db  11974  ACAGAAGTATTTGAGTAGTAGTATTTGAAAAAGAGAATTTTTTAATGGTATACATTCAAAAC  11915

Qy    572  --------
Db  11914  TACTAATTTTTCATCAACATCACTTCGTATCAACCCATGACCAAAGTCTCGTTATTATA  11855
```

```
DB:                          3                    Gaps:              2

US-10-736-868-2 (1-643)  x  CBRG33D04 (1-39089)

Qy      1 MetIleLeuPheLeuPheLeuPheLeuLeuLeuGlyLeuPheCys---------------- 14
          |||::::::|:||||::|||:||  :::|:||:||||    :::||||
Db   1805 ATGTTAATATTTTTACTAATTTTTCTGTTTCTCCAATTTTGTAAGTTTTTACAGTTCAA 1746

Qy     15 --------------------------------------------IleAl 16
                                                          :::
Db   1745 AAGTTGGTCAATGGAAGGAAGAACTTCGGTTACAATCATTTCTTTTAACTTTCTAGA 1686

Qy     16 aProLeuSerAlaGlnSerProSerThrSerAspAlaProGlyAlaLeuLeuSerSerLe 36
          |||::||||||||  ||||||    ||||||||  |||||||||
Db   1685 TCCAATATCCGGGGCTCACCACCGAATTCCGATGTTCCCGGTGCTCTTTATCCTCTCT 1626

Qy     36 uValGlyLysSerHisGlnLysProLeuAlaProSerMetGluAlaLeuGluLeuMe 56
          |||:::||||||   |||||  |||||||
Db   1625 TGCCACTAAAAACCACCCAAAACTTCCATTAGCTCCATCCATGGAAGCACTAGAATTGAT 1566

Qy     56 tGlyValGlnPheValAspAlaLeuuIleLysLysGlyGlnMetAlaLysGlyAl 76
          |||||||:||||||||||||||||  |||:::||||||
Db   1565 GGGAGTTCAATTTGTGGATGCTTTAATTAAAAAAGGACAGACAGATGAAATGCAAAGGAGC 1506

Qy     76 aPheLysThrGlnLeuGluValGluLysValHisProAspGlnPheAspLysTyrLy 96
          |||||||||||||||  ||||||||  ||||||||||::|||||||
Db   1505 ATTCAAAACAACTGGAGGTTCTGGAGAAGTACATCCGGATCAATTTGAGAAATATAA 1446

Qy     96 sLysLeuValAspAspLeuAlaLeuAspAlaValMetGlnGlnAlaGluMetAlaLy 116
          ||:::||||||||||||  ||||||||||||  :::|||||
Db   1445 GAAACTGAAAGTGGAGGATTTAGCCGCAGATGCCAGATTATCCAACAGGCAGAAATGTCCAA 1386

Qy    116 sLeuGlnProLys 120
          |||||||||||
Db   1385 GTTGCAACCGAAAACTGGTAAGGAGAAAATATAAAAGAGAGATCATTACTACGAGCAATA 1326

Qy    120 ------------------------------------------------------------ 120
Db   1325 CGACTGAAAATTAATTTCTAGTTAGCACATGCAACAGTAAAAAATTCAAAATTCAGTTTT 1266

Qy    120 ------------------------------------------------------------ 120
Db   1265 GAGATAATGTCAGAAAAAAGTTTTTATTTTCACATTTTAATAAATGTAAGTATCAATTC 1206

Qy    120 ------------------------------------------------------------ 120
Db   1205 CGATCGAATCAATTTGGCAAATCATTAGAGGAATAAGTTTCTGGAAACTGAAAAGTTTT 1146

Qy    120 ------------------------------------------------------------ 120
Db   1145 CAAATGTCAGATCTGAAACTGAACTTTCGAAAGTCTGAAAAACCTTAAGTAAATAAACATT 1086

Qy    120 ------------------------------------------------------------ 120
Db   1085 GAACACTGTCTAATAGACAGGGCGTCCTAACTTCAAGTTTCGAGACTTTTATATTTTTGG 1026

Qy    120 ------------------------------------------------------------ 120
Db   1025 TGAAAAATGAGTAATAGTTCCATCCGGGAACTTTAGTTCCGACTTTCAAGTTTAAAAATT 966

Qy    120 ------------------------------------------------------------ 120
Db    965 CGTCTTCCCAAAACTTTAAATCGTACAAGCAGAAATTGTGATGATGAAGACCAAGAAGTG 906

Qy    120 ------------------------------------------------------------ 120
Db    905 ACACTTACAATATATAAAAACCTAATATAACGCCCGAAACAGTTATTTTCTAGTCGACT 846

Qy    121 --------SerGlyAsnAlaPheLeuAspMetLeuAsnGlyAsnGlyIleProIleGlySe 138
                  |||:::|||||||
Db    845 ATCTTTTTCAGGAAACGCATTCATCGATATGTTAAATGGAAATGGAAATTCCAATTGGAAG 786

Qy    138 rSerIleArgGlyLeuGluAspAlaIleIleArgThrGlnArgAspMetGluAsnThrAspPr 158
```

```
RESULT 4
CBRG33D04/c
LOCUS       CBRG33D04              39089 bp    DNA     linear   INV 04-NOV-2000
DEFINITION  Caenorhabditis briggsae cosmid G33D04, complete sequence.
ACCESSION   AC084550
VERSION     AC084550.1  GI:11095000
KEYWORDS    HTG.
SOURCE      Caenorhabditis briggsae
  ORGANISM  Caenorhabditis briggsae
            Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
            Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE   1  (bases 1 to 39089)
  AUTHORS   Washington University Genome Sequencing Center
  TITLE     The C. briggsae Genome Sequencing Project
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 39089)
  AUTHORS   Wilson,R.
  TITLE     The sequence of C. briggsae cosmid G33D04
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 39089)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-NOV-2000) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT     Submitted by:
            Genome Sequencing Center
            Department of Genetics, Washington University,
            St. Louis, MO 63110, USA
            e-mail: jspieth@watson.wustl.edu

            NOTICE: This sequence may not be the entire insert of this clone.
            It may be shorter because we only sequence overlapping sections
            once, or longer because we provide a small overlap between
            neighboring submissions.
FEATURES             Location/Qualifiers
     source          1..39089
                     /organism="Caenorhabditis briggsae"
                     /mol_type="genomic DNA"
                     /strain="GuJArat G16"
                     /db_xref="taxon:6238"
                     /clone="G33D04"
ORIGIN
```

```
Alignment Scores:
Pred. No.:              2.58e-27       Length:               39089
Score:                  585.00         Matches:                141
Percent Similarity:     41.37%         Conservative:            10
Best Local Similarity:  38.63%         Mismatches:              11
Query Match:            17.78%         Indels:                 203
```

```
Db   785 TAGTATTCGTGGATTAGAAGATGCTATTCGGACACAAAGAGACATGAGAACACTGATCC 726
         ||||||||||||||||||||||||||||||||||||||||||||||||||

Qy   158 oSerGluGlnIle 162
         ||||||||||:::
Db   725 ATCTGAGCAGGTA 713
```

RESULT 5
CQ575340

```
LOCUS       CQ575340                3631 bp    DNA        linear      PAT 02-FEB-2004
DEFINITION  Sequence 3098 from Patent WO0171042.
ACCESSION   CQ575340
VERSION     CQ575340.1  GI:41638873
KEYWORDS    .
SOURCE      Drosophila sp.
  ORGANISM  Drosophila sp.
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1
  AUTHORS   Venter,J.C.; Adams,M., Li,P.W. and Myers,E.W.
  TITLE     Detection kits, such as nucleic acid arrays, for detecting the
            expression of 10,000 or more Drosophila genes and uses thereof
  JOURNAL   Patent: WO 0171042-A 3098 27-SEP-2001;
            PE Corporation (NY) (US)
FEATURES             Location/Qualifiers
     source          1..3631
                     /organism="Drosophila sp."
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:7242"
ORIGIN

Alignment Scores:
Pred. No.:              0.0268
Score:                181.00
Percent Similarity:    35.51%           Length:         3631
                                        Matches:         165
Best Local Similarity: 21.94%           Conservative:    102
Query Match:            5.50%           Mismatches:      258
                                        Indels:          228
DB:                                     Gaps:             37
```

US-10-736-868-2 (1-643) x CQ575340 (1-3631)

```
Qy    39 LysSerHisGlnLeuProLeu--------------------------------AlaProSer 49
         |||||:::||:||||
Db  1031 AAAGACAACAAGAAAGATCCACTGATTGCGAGGATTCCAAACCAAAGAGAGTCCCTAAG 1090

Qy    50 MetGluAlaLeuGluLeuMetGlyValGlnPheValAspAlaLeuIleLysLysGlyGln 69
         |||                     :::   |||||
Db  1091 AAGGAAACAACAGCAAAACCGGAGA------------------------AAAGGCAAG 1123

Qy    70 MetGluMetAlaLysAlaLysGlyAlaPheLysThrLeuGlnGluValLeuGluValLeuGlnGluGlnLysProLysSerValHisPro 89
         |||                     |||||:::::||
Db  1124 AAGGAGGAGGAGCAAACCAGCTGAAAAGAGTAAGAAAAGAGAAGAAGGAA------ 1171

Qy    90 AspGlnPheAspLysTyrLysLysLeuGlyValAspAspLeuValAspAspLeuAlaAlaAspAlaValMet 109
         ||||||||                   ||| :::
Db  1172 GACAGCACAACAAGTCAAGAAGAAGAAAGCAGAACCAGAAAATAATCAAAAAGAA 1231

Qy   110 GlnGln 111
         |||||
Db  1232 CAACAAATTGGAGTCAAGAAAAAAGACTTCTGAACCAGAAGATGCATCGAACTCACACAAA 1291

Qy   112 AlaGluMetAlaLysAlaLysLeuGlnProLysSerGlyAsnAlaPheLeuAspMetLeuAsnGly 131
         :::                     |||||
Db  1292 TCGGTGACGGCCAAAAAGAGGATTCCAAGAAAAGACGAGGCATCTACTCACTTAAGTCGGAC 1351

Qy   132 AsnGlyIleProIleGlySerSerIleArgGlyLeu------ 143
         |||                     :::
Db  1352 AACGAATCTCCGGAAGCAGACAAATCTGATGATCCTCTCGGAGGCGAGGCCAACAGT 1411

Qy   144 --------GluAspAlaIleArgThrArgAspMetGluAsnThrAsp---ProSer 159
                 |||                 |||:::
Db  1412 TCCAGCAGCGAGCGACGAGATGCGGATACGACACCGCCGACACCAGACCCTCC 1471
```

```
Qy   160 GluGlnIleAlaLysAlaValMetAspLysPheGlnThrGlnIleLeuProGlyLeuVal 179
         :::                     |||
Db  1472 GCCCAA----------------------------AAGGAGACTCGCCTGAACAGGAGAGTCGCCGAA--GTA 1516

Qy   180 AlaAsnMetIleAlaGlyLysAsnPro------ 197
         ||||:::                 |||:::||
Db  1517 ACCGGCGCCCCGAAGGTCTTACGCCCAAACAACGACACGACTCATGAGCAAAGGAAAAAG 1576

Qy   198 AlaGlnAlaAlaProSerSerValPheGlnIleAlaLeuGlnAlaArgAlaMetLeuGly 217
         |||||:::                :::|||||||
Db  1577 GCTCGCCAGGAGAAG------GAACAAAGCTGGCCGAGGAGCGTCGCCTCAAG 1624

Qy   218 LysAsnAlaProValAlaGlyGlyArgGlyGluGlnArgMetMetAsnArgVal 237
         |||||                   |||||||
Db  1625 CAACAG------GATAAGGAGCACCGCAGCAAAAGAAACAGAGCGC-- 1669

Qy   238 AspGlnArgMetGlnArgGluLeuGln--GluGluAspGluAspAspAspLeu 256
         |||||:::                :::  :::
Db  1670 GACGAAAAGGAACAGCACGCGGAAGCTGGAGGAGGGATCAGAAGGAGCAGCAGGAAGAATG 1729

Qy   257 GluAspGluAspValProArgArgArgSerArgSerAspGlyGlnProGlnSerGlnAlaGlu 276
         ||| |||||               :::::||
Db  1730 GAAAAGGAGGAGGAAGGAGCCAAGCAACGACAGCGGCGAGTGGGACTCCAGAACGAGGAGAAG 1789

Qy   277 HisGlnArg----------ArgAspLeuAlaArgArgLeuLysSerSerProArgLeuLys 293
         ::||                     :::||
Db  1790 CGCAAGCGCAACGAGGCCAGGAGGAGGTGCAACGCAAAAGGAGGAGGAGCGTCGCAAG 1849

Qy   294 GluLeuLeuGlnLysAsnAlaGluValGln-------------- 302
         :::                     :::
Db  1850 AAGGAGCAGGAGCGGTGAGGAGGCCGAGCAAAAGAAGAAGAGCGCGCCGCCGAGTCCTTCTCC 1909

Qy   303 ------------------------------------SerLeuLeu 305
                                             |||
Db  1910 AAGTTCTTCGTGCCCAAGCACATGCGGCGCCGGATCCAACAACACATCCTACCTG 1969

Qy   306 SerTyrGlnArgMetArg---AspSerProLeuSerLysArgArgProLeuAsnLeuIleAspProSer 324
         :::::::                 :::
Db  1970 GAGCACGCCCAAGCACCTGCGACAGCTCCAAGGCCACCAGCAGCAAACGCTG---------- 2020

Qy   325 AspGluAspGluSerAlaPheArgAlaMetGluAlaArgAlaAlaLysLeu 340
         :::                     |||
Db  2021 --------------------GCCTTTCGACCATTTCAGATCAAGGACGACATGCTGCTCGACCACCG 2065

Qy   341 ----------------AspGlnLysSerGlnLeu--ValLeuGlyLeuHis 351
                         ::::||||||               |||
Db  2066 ATCGTCGCAACTGCGTCCGCGGCCGCGAGCTCCCAGCTAGATGGACTTTTTCGACAC 2125

Qy   352 GlyPheGlySerAspAspAspGluAspGluAsnLeuIleAspProProSer 371
         |||:::::                ||||||
Db  2126 CGGGACGAGGAAGCTGCGACGACGAGGAAGAGGAGGAGGAGCAGCAGCATCGTC- 2176

Qy   372 GluAsnSerPheArgArgAlaPro----------------------LeuArgLeuSer 383
         :::                     |||
Db  2177 ------------------AGAGGAAAGCCGCCGAATCGGGCTAATCTCTATCTTAGTGAGTTGAGC 2224

Qy   384 SerGlyPheValGluLysSerAlaAsnGluGluLysSerAlaLeuAspArgIle 403
         :::::::                 |||
Db  2225 AGCGAGAGCCCAAGCGCCGCTCAAGATGCAGCCGATGTGAAG---CTCCAAAGACGCACC 2281

Qy   404 LysTyrArgValAspAsp-------------------ValGluLysTyrLeuAlaPro 416
         |||                     :::::
Db  2282 AAGGACAGCAGGAGGATGATGACGATGTTCAAGTAGATGATTACCTTTCACCGGGGTCTG 2341

Qy   417 ----------------------------LysProMetGluPhe 421
                                     |||
Db  2342 CCCATCGAGGTGGAGCAGCGCCAAGCAGCTGACCCGAATGAGGCCAAGTATCTGCATTTC 2401

Qy   422 AsnProLysProGlnProGlyTyrPhe---Ala 431
         |||                     :::::||
Db  2402 GCCGACAACCGGAGCCCACCTACTACGGACCACCTGCGTAAAAAGAGCAGCATCTCC 2461
```

```
University of Colorado Health Sciences Center, 4200 East Ninth Ave,
Denver, CO 80262, USA
```

FEATURES   Location/Qualifiers
     source   1..3745
               /organism="Drosophila melanogaster"
               /mol_type="mRNA"
               /db_xref="taxon:7227"
     CDS   42..3593
               /codon_start=1
               /note="dCAF-1 p180"
               /product="chromatin assembly factor-1 p180 subunit"
               /protein_id="AAK31263.1"
               /db_xref="GI:13569827"
               /translation="MHAGVKTPLSGRKKDAAVPAKSAEKTASGGKKFVQTRLPFKLL
               TPGGGAVPTSSSSSSPATIGSGPVTVILDEDDPAPRKRKLSYDDESPSEGTGCSTGQ
               LRRSTSKENLDLASSIATKKVKTTDSVVEDVIELDEDBADKEIEDDQLVEAKSSKEV
               KLKPKKSEGKKGSPAPIQIKLPLVNKSKRRSLKKSEEPLETSTADVAKGDSESSDD
               IEIIAEELNPQKRQKVQTPREKSPENSSTSGKHTEEVMQKNGKDSKKQEESMEAVKES
               KKEEPKTAEQNKKNDQATIDLFMGKKMETNKKDKPEVKTTAKDNKKDPLIAEDSKPKE
               VPKKETSKTGGKGKEGAKPAEKSKKEEKEDSTTKSKKEADSPANNOKEQQIGVKKK
               TSEPEDASNSHKSVTAKKDSKKDEASTQVKSDNESPEAAEISMILSTSEANSSSEHE
               MDADTDTAATDRPSAQKETRLNRKSLPEVTGAPKGLTPKQQRLMEQRKKAREEKEQKL
               AEBRRLKQQDKEHREQQKKQERDEKEQQRKLERDQKSEQQRKMEKEEKERKRQAEVDSK
               NEEKRKRNEAKEEVQRKKDEERRKKEQEREEAKEGVRQAAESFSKFVPKQPKKCGSGS
               NNTSYLEHEQSSCDSSKASSQTIAFRPFPQIKDDMLLAPIVRNSLGEQORSQLDGLFRH
               RDEEADDEEEEEQDIVRRKPNRANLYLSELSSGRKPLKMQRDVKLQRRTKDEEDD
               DDVQVIDYLSPAGLPIEVEQPKQLTRMKAYLHFADNRRPPYYGTMRKKSSSISARRP
               LAQDKVLFDYEVDSDCEWEEEEPGESLSASEDEKEKESEEESEENEWYVPHGHLSD
               EELQNDGDMEDGHTREAQKAKLQVLQQEFAQEMKQTKKIKARLLGPVWLDENGNKS
               ELFPAIFAHTIDMYACWQVEPLSLEPPPEPERQDQTPEQPVPLQLDDRLMQQLVRLTH
               GNRNSKTFLINEYLEYLKTQATAEANQTTLPSKAVLREKFDELASWKTVELSTPEAAA
               ASANSAKKGRKPKKRLCWVVASDMLEKFQLPDLQLNQWNYLTPKVSDGPDAAQQEQ
               SPPADAVPAASPASNQTTNSAFTTPISGQTQAGQGVKKRATLIMSVPRDQPIPSATKN
               KLISQFLQRQSDASKPKEKMQEKPPPPIVFDDVVMLSD"
```

ORIGIN

```
Alignment Scores:
Pred. No.:               0.0279          Length:        3745
Score:                 181.00           Matches:        165
Percent Similarity:     35.51%          Conservative:   102
Best Local Similarity:  21.94%          Mismatches:     258
Query Match:             5.50%          Indels:         228
DB:                      3              Gaps:            37
```

```
US-10-736-868-2 (1-643) x AF367177 (1-3745)
```

```
Qy   39 LysSerHisGlnLysProLeu-----------------------------------AlaProSer 49
           ||||  ||||:||
Db  993 AAAGACAACAAGAAGAATCCACTGATTGCAGAGGATTCCAAACCAAAAGAAGTCCTAAG 1052

Qy   50 MetGluAlaLeuGluLeuMetGlyValGlnPheValAspAlaLeuIleLysLysGlyGln 69
           |||  :||||| ||||||||    |||
Db 1053 AAGGAAAGCAAAACCGGAGA-----------------------------AAAGGCAAG 1085

Qy   70 MetGluMetAlaLysGlyAlaPheLysThrGlnLeuGluLeuLysLysValHisPro 89
           |||||||||:|||  ||||||||||:|||||
Db 1086 AAGGAGGAGCAAAACCAGTGAAAAGAGTAAGAAAGAAGAGAGGAA--------- 1133

Qy   90 AspGlnPheAspLysTyrLysLysLeuValAlaAspAspLeuAlaAlaAspAlaValMet 109
           ||||||  ||||::
Db 1134 GACAGCACAACAAGTCAAAGAAAGAGAAAGCAGACTCACCTGCAAATAATCAAAAAGAA 1193

Qy  110 GlnGln------------------------------------------------- 111
           ||||||
Db 1194 CAACAAATTGGAGTCAAGAAAAAGACTTCTGAACCAGAAGATGCATCGAACTCACACAA 1253

Qy  112 AlaGluMetAlaLysGlnProLysSerGlyAlaPheIleLysAlaPheIleAspMetLeuAsnGly 131
           :::    |||||||  |||
Db 1254 TCGGTGACGGCCAAAAGAGATTCCAAGAAAGACCAGGCATCTACTCAAGTTAAGTCGGAC 1313

Qy  132 AsnGlyIleProIleGlySerIleArgGlyLeu------------------------------ 143
           :::      ||||||
Db 1314 AACAATCTCCGAAGCAGCAGAAATCTCGATGATCCTTCTGACCAGGCAGCGGAGGCCAACAGT 1373
```

```
Qy  432 ProArgLysIleProThrArgProArgLysMetLeuProLeuLeuIleGlySerAspPro 451
           |||   :::   :::         ||||||
Db 2462 GCCCGACGGCCCGCTGCACAGGATAAGGTGCTCTTCGACTACGAAGTGGACTCCGATTGC 2521

Qy  452 LysValGlnGluGlnLeuArgArgHisPro--------------SerThrGluTrpLys 466
           ::: ||||||||| ||||              ::: |||
Db 2522 GAGTGGGACGGAGAG----------GAGCCCGGGGAATCGTTGAGCGCTAGCGAGGACGAG 2572

Qy  467 IleAlaLysGluSerArgValLeuThrArgLeuLysAsnProSerLeuAla------- 484
           ||||||                  ::: |||
Db 2573 AAAGAGAGGAGTCCGAGGAGGAATCGGAAGAGGAGTACAACGAGTGGTACGTGCCGCAC 2632

Qy  485 AlaLeuPheMetAspAspLysGluGlnLeuAsnThrLeuLysGlyArg----------- 499
           ||||:::|||||||||||| :::
Db 2633 GGACATCTCAGCGACGACAGCAACTGCAAAACACGGCGGAATGGAAGACGGCCACACG 2692

Qy  500 ---------------GlnMetLeuThrAspGlu------GlnLysGlyArg 509
                          ||||||||
Db 2693 CGGGAGGGGCAGAAGCCAAACTCCAGGTGCTGCAGCAGGAGTTCGCTTCAGGAAATGAAG 2752

Qy  510 ThrArgValLysThrIleArgAlaLeuProArgLeuPheGlyAla-------------- 524
           :::       |||||||||:||::|||
Db 2753 AAGCAGACCAGACAGCAAAAAGATCAAGGCG------CGCCTGTGGTGGCCGTTGGCTGGACGAG 2806

Qy  525 ------------------ProThrAlaLysAlaGluMetIleAspAlaAlaLysVal 536
                              ||||||           |||:||
Db 2807 AACGGCACACAAGTCCAGCTATTCCAGCTATCTTGCCACACCATGCACATGTACGCC 2866

Qy  537 PheGlnAspIleGluArgProIleProLeuPheGluProLysProLysGlyArgHis 556
           ||||||                     |||||||
Db 2867 TGTTGGCAGGTGGAG----------CCACTCAGCCTAGAGCCTCC-GCCAGA---- 2907

Qy  557 ThrArgLeuArgTrpThrArgGlyAlaAsnGluLysGlyLeuProGlyLeuGlySerArgPhe 576
           ||::::           :::    :::
Db 2908 ---------------ACCGGAGGTCAGGATCAGACGCCAGAGCAGCCGGTGCCGCTGCA 2952

Qy  577 IleLeuProSerLeuAspProThrMetProAlaLeuAsnThrAlaPheSerThrGlnGly 596
           |||        :::    |||
Db 2953 GCTGGACGATCG-----------TCTCATGCAGCTGCGTGGCGCCTTACCCACCACGGCAATCG 3003

Qy  597 ArgAlaArgAspGluTrpThrAsp---ThrMetPheLysThrProAlaAsnTrpAsnProGly 615
           |||                  ||| ::: |||
Db 3004 GAACTCGAAGACGTTCCTGATCAACGAATATTTAGAGTACCTCGAA--------AACCCAAGC 3057

Qy  616 AspGluValGlyPheLysMetAsnSerLysThrLys 627
           |||                ||||||
Db 3058 GACCGGCGAGGCCAATCAAACACACGCTGCCCTCAAA 3093
```

```
RESULT 6
AF367177
LOCUS       AF367177               3745 bp    mRNA    linear    INV 19-SEP-2001
DEFINITION  Drosophila melanogaster chromatin assembly factor-1 p180 subunit
            mRNA, complete cds.
ACCESSION   AF367177
VERSION     AF367177.1    GI:13569826
KEYWORDS    .
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 3745)
  AUTHORS   Tyler,J.K., Collins,K.A., Prasad-Sinha,J., Amiott,E., Bulger,M.,
            Harte,P.J., Kobayashi,R. and Kadonaga,J.T.
  TITLE     Interaction between the Drosophila CAF-1 and ASF1 chromatin
            assembly factors
  JOURNAL   Mol. Cell. Biol. 21 (19), 6574-6584 (2001)
  MEDLINE   21424640
   PUBMED   11533245
REFERENCE   2  (bases 1 to 3745)
  AUTHORS   Tyler,J., Kadonaga,J. and Kobayashi,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-MAR-2001) Biochemistry and Molecular Genetics,
```

```
Qy  144 --------GluAspAlaAlaIleArgThrGlnArgAspMetGluAsnThrAsp---ProSer 159
             |||       :::                 ||||||  |||
Db 1374 TCCAGCAGCGACGAGCACGAGACGATGCGGATACGGACCACCGTACTACGGCACCGACACCCTCC 1433

Qy  160 GluGlnIleAlaLysAlaValMetAspLysPheGlnThrGlnIleLeuProGlyLeuVal 179
             :::       ||||||:::||||||:::         |||
Db 1434 GCCCAA----------AAGGAGACTCGCCTGAACACAGGAAGAGTCTGCCGGAA---GTA 1478

Qy  180 AlaAsnMetIleAlaLeuGlyLysAsnPro------PheLysMetProGlnMetArgLys 197
        |||         |||         :::::::|||||| |||
Db 1479 ACCGGGGCCCCGAAGGGTCTTACGCCCAAGACGACTCATGAGCAAGGACAAAGGAAAAG 1538

Qy  198 AlaGlnAlaAlaProSerSerValPheGlnIleGlnAlaLeuAlaGlnAlaMetLeuGly 217
        :::||| ||||||:::
Db 1539 GCTCGCGAGGAGAAG--------GAACAAAAGCTGGCGAGGAGGAGAGAGCTGCGCCTCAAG 1586

Qy  218 LysAsnAlaProValAlaGlyIleGlyArgGlyGluGlnGlnArgMetMetAsnArgVal 237
        :::
Db 1587 CAACAG--------GATAAGAGAGCATCGCGAGCAGCAAAAGAAACAGGAGCGC-- 1631

Qy  238 AspGlnArgMetGlnIleArgGlnLeuGln---GluGluAspAspAspAspAspLeu 256
        ||||||:::||||||:::     ||||||  :::       |||
Db 1632 GACGAAAAGGAACAGCAGCCGAAGGCGTGGAGAGGGGATCAGAAGGGAGCAGCAGAGAAAGATG 1691

Qy  257 GluAspAspGluValProValProArgArgSerArgAspGlyGluProGlnSerGluAlaGlu 276
        |||||| ||||||     :::     ||||||
Db 1692 GAAAAGGAGGAGAAGGAAGGAGTTGGAGCCGCAGCGGAGGGTGGAGACTCCAAGAACGAGGAGAAG 1751

Qy  277 HisGlnArg----------ArgAspLeuAlaArgArgLeuLysSerSerProArgLeuLys 293
        :::||| |||:::
Db 1752 CGCAAGCGCCAACGAGCCAAGGAGTGCAACGCAAAAAGGACCGAGGAGCGCCCCAAG 1811

Qy  294 GluLeuLeuLeuAsnAlaGluValGln-------------------- 302
        :::          :::
Db 1812 AAGGAGGAGGAGGCGTGAGGAGGCCGAGCCAGAGAAGAAGAAGCGCGCCGAGTCCTTCTCC 1871

Qy  303 --------------------------SerLeuLeu 305
                                  |||
Db 1872 AAGTTCTTCGTGCCCAAGTCGCCGAGCCAGGGATCGCAGCCAACACCATCCTACCTG 1931

Qy  306 SerTyrGlnArgMetArg----AspSerLeuSerLysArgArgArgProLeuAlaMetAsn 324
        :::          ||| :::
Db 1932 GAGCACGACAAAGCCGCACAGCTCCAAGGCCCAGCAGCCAAACGCTG------------ 1982

Qy  325 AspGluAspGluSerAlaPheArgAlaMetGluAlaMetGluAlaLysLeu--- 340
        |||||| :::             :::       :::
Db 1983 --------GCCTTTCGACCATTTCAGATCAAGGACGACATGCTGCTGGCACCG 2027

Qy  341 --------AspGlnLysSerGlnLeu---ValLeuGlyLeuHis 351
                :::||||||||||     ||| :::||||||
Db 2028 ATCGTTGCGCAACTCGCTCGGCGGCAGCAGGCCCCCAGCTAGATGGACTTTTTCGACAC 2087

Qy  352 GlyPheGlyGluSerAspAspAspGluGluAspGluAspAspGluAsnLeuIleLeuAspProSer 371
        |||||||||:::||||||:::||||||
Db 2088 CGGGACGAGGAAGCTGACGACGAGGAAGAGGAGGAGGAGCATCGTC-------------- 2138

Qy  372 GluAsnSerPheArgArgAlaPro-------- 383
        ||||||
Db 2139 --------AGAGCGAAAGCCGCCCAATGCGGGCCTAATCTATCTTAGTGAGTTGAGC 2186

Qy  384 SerGlyPheGlyValGluLysLeuLeuLysSerAsnAspGluLeuLysSerAlaLeuAspAspArgIle 403
        |||||| ::::::     :::
Db 2187 AGCGGACGCCGGCCAAGACGCCGTTCCTCAAGATGTCGAGG---CTCCCAGAGACGCACC 2243

Qy  404 LysTyrArgValAspAsp------------ValGluLysTyrLeuAlaPro 416
        :::          |||||||||:::
Db 2244 AAGGACGAGGAGGATGATGACCGATGTTCAAGTGTTCAAGATTACCTTTCGCCCGGGGACTCG 2303

Qy  417 --------LysProMetGluPhe 421
                :::    |||
Db 2304 CCCATCGAGGTGGAGCAGCCCAAGCAGCTGACGACGAATGAAGGCCAAGTATATCTGCATTTC 2363
```

```
Qy  422 AsnProLysProGlnProGlyTyrPhe-----------Ala 431
            :::||||||:::                      :::
Db 2364 GCCGACAACCGGAGGCCACCGTACTACGGCACCGACACCCTCGGCAAAAGAGCAGCAGCATCTCC 2423

Qy  432 ProArgLysIleProThrArgProArgLysMetProLeuIleGlySerAspPro 451
            :::        ||| :::|||:::
Db 2424 GCCCGACGCCGCTGGCACAGGATAAGGTGCTTCGACTACGAAGTGGACTCCGATTGC 2483

Qy  452 LysValGlnGluGlnIleLeuArgArgHisPro------SerThrGluTrpLys 466
        :::          ||||||:::            ||||||:::
Db 2484 GAGTGGAGGAGGAG--------GAGCCGGGGGAATCGTTGAGCGGCTAGCGAGGACGAG 2534

Qy  467 IleAlaLysArgValLeuValLeuThrArgLeuLysAsnAsnProSerLeuAla------ 484
        |||        :::          ||||||
Db 2535 AAAGAGAAGGAGTCAGAGGAGGAATCGGAAGAGGAGTACAACCGAGTGGTACGTGCCCCAC 2594

Qy  485 AlaLeuPheMetAspAspLysLeuGluAsnThrLeuLysGlyArg--------- 499
        |||:::|||     |||:::
Db 2595 GGACATCTTAGCGACGAGGAACTGCAAAACGACGGCCACGGAGTTCGCTCAGGAAATGAAG 2654

Qy  500 --------GlnMetLeuThrArgAspGlu------ 509
                |||||||||:::
Db 2655 CGGGAGGCGCAGAAGGCCAAACTCCAGGTGCTGCAGCAGGAGTTGGCTGGACGAG 2714

Qy  510 ThrArgValValLysThrIleArgIleLeuProArgLeuPheGlyAla----- 524
        :::                    ||||||
Db 2715 AAGCAGCAGAAAAAGATCAAGGCG--------CGCCTGTTGGGTCCCGTTTGGCTGGACGAG 2768

Qy  525 --------ProThrArgLeuAlaGluMetIleAspAlaLeuLysVal 536
                |||||||||:::
Db 2769 AACGGCCAACAAGTCCGAGCTATTCCCACCTATCTTTGCCCACCATGTACGACGCC 2828

Qy  537 PheGlnAspIleGluArgProIleProLeuPhePheGluProLysGlyArgHis 556
        |||||||||:::          ||||||
Db 2829 TGTTGGCAGGTGGAG--------CCACTCAGCGCTAGAGCCTCC--GCCAGA-- 2869

Qy  557 ThrArgLeuArgTrpThrGlyValAlaAsnGluLysGluIleProGlyLeuGluSerArgPhe 576
                ||||||:::       ||||||:::
Db 2870 --------ACCGGAGCGTCAGGATCAGAGCCGCAGCAGCCGGTGCCGCTGCA 2914

Qy  577 IleLeuProSerLeuAspProThrMetProAlaLeuAsnThrAlaPheSerThrGlnGly 596
        |||:::       ||| :::
Db 2915 GCTGGACGATCG--------TCTCATGCAGCAGCTGGTGCCCTTACCCACCGGCAATCG 2965

Qy  597 ArgAlaArgAspGluTrpAsp---ThrMetPheLysIleProAsnAsnTrpAsnProGly 615
        |||       :::::::|||::::::
Db 2966 GAACTCGAAGACGTTCCTGATCAACGAATATTTAGAGTAGCCTGAA--------AACCCAAGC 3019

Qy  616 AspGluValGlyPheLysMetAsnSerLysThrLys 627
        |||       :::
Db 3020 GACCGCCGAGGCCAATCAAACAACAACGCTGCCCTCAA 3055
```

```
RESULT 7
ACMHC
LOCUS          ACMHC              5894 bp    DNA      linear    INV 10-FEB-1999
DEFINITION     Acanthamoeba castellanii nonmuscle myosin heavy chain gene.
ACCESSION      Y00624 M12702 M12703 M19549
VERSION        Y00624.1 GI:5585
KEYWORDS       myosin heavy chain.
SOURCE         Acanthamoeba castellanii
  ORGANISM     Acanthamoeba castellanii
               Eukaryota; Acanthamoebidae; Acanthamoeba.
REFERENCE      1
  AUTHORS      Hammer,J.A. III, Bowers,B., Paterson,B.M. and Korn,E.D.
  TITLE        Complete nucleotide sequence and deduced polypeptide sequence of a
               nonmuscle myosin heavy chain gene from Acanthamoeba: evidence of a
               hinge in the rodlike tail
  JOURNAL      J. Cell Biol. 105 (2), 913-925 (1987)
  MEDLINE      87308395
  PUBMED       3040773
REFERENCE      2  (bases 1 to 5894)
  AUTHORS      HammerIII,J.A.
  TITLE        Direct Submission
```

```
JOURNAL   Submitted (19-OCT-1987) HammerIII J.A., Lab. of Cell Biology, BLD6
          3, RM B1-22, National Institutes of Health, Bethesda, MD 20892
FEATURES          Location/Qualifiers
     source        1..5894
                   /organism="Acanthamoeba castellanii"
                   /mol_type="genomic DNA"
                   /db_xref="taxon:5755"
                   /clone_lib="lambda 2001"
     precursor_RNA 517..>517
                   /note="minor primary transcript"
     mRNA          join(560..895,1161..1310,1381..1565,1743..>1743)
     precursor_RNA 560..>560
                   /note="major primary transcript"
     CDS           join(689..895,1161..1310,1381..1565,1743..5730)
                   /codon_start=1
                   /product="myosin heavy chain"
                   /protein_id="CAA68663.1"
                   /db_xref="GI:5586"
                   /db_xref="GOA:P05659"
                   /db_xref="UniProt/Swiss-Prot:P05659"
                   /translation="MAAQRRRKGGEVESDYIKYLKYKNTGFQVSASDKTLAWWPTKDA
          DRAFCHVEVTKDDGKNFTVRLENGEEKSQPKNEKNFLGVNPPKFDGVEDMGELGYLNE
          PAVLHNLKKRYDADLFHTYSGLFLVVVNPYKRLPVTTPEIIDIYRGRQRDKVAPHIFA
          ISDAAYRAMLNTRQNQSMLITGESGAGKTENTKKVIQYLTAIAGRABGGLLEQQLLEF
          NPILEAFGNAKTTKNNNSSRFGKFIELQFNAGGQITGANTFIYLLEKSRVTAQGAGER
          NFHIFYQILSKAMPEELKQKLKLJTKPEDYFFLNQNACYTVDDMDDAKEFDHMLKAPDI
          LNINEEERLAIFQTISAILHLGNLPFIDVNSETAGLKDEVELNIAAELLGVSAAGLKA
          GLLSPRIKAGNEWVTRALNKPKAMASRDALCKALFGRLFLWIVQKLNRILSHKDKTAL
          WIGVLDISGFEIFQHNSFEQLCINYTNEKLQQFFNHHMFTLEQQEYEREKIDWTFVDY
          GMDSQDCIDLIEKKPMGILPLLDEQTVFPDADDTSFTKKLFQTHENHRNRPRRPFDAN
          NFKIVHYAGEVEYQTSAWLEKNRDPLEDDLSNLCKKSSVRFVTGLFDEDLMPSFKAAP
          AEEEKAAAGGSSNNRSTGRGKGGAQFITVAFQYKEQLAHLMSMLSSTAPHFIRCIIPNL
          GKKPGVVSDQLVLDQLKCNGVLEGRIRARKGWPNRLKYDEFLKRYFLLKPGATPTSPS
          TKDAVKDLIEHLIAKEPTKVNKDEVRFRGVTKIFFRSGQLAAIEELREQAISKMVVSIQ
          AGARAFLARRMYDKMREQTVSAKILQRNIRAWLELKNWAWYQLYVKARPLISQRNFQK
          EIDDLKKQVKDLEKELAALKDANAKLDKEQLAEEDADKLEKDLKALKLKILDLEGEK
          ADLEEDNALLQKKVAGLEEELQEETSASNDILEQKRKLEAEKGELKASLEEERNRKA
          LQEAKTKVESERNELQDKYEDRAAAHDSLKKKEEDLSRELRETKDALADAENISETLR
          SKLKNTERGADDVRNELDDVTATKLQLEKTKSLEEELAQTRAQLEEEKSGKEAASSK
          AKQLGQQLEDARSEVDSLKSKLSKSLSAABKSLKTAKDQNRDLEQLEDERTVRANVDKQKK
          ALEAKLTBLEDQVTALDGQKNAAAAQAKTLKTQVDETKRLEEABASAARLEKERKNA
          LDEVAQLTADLDAERDSGAQORRKLNTRISELQSELENAPKTGASSEEVKLEGELE
          RLEEELLTAQERAAAAEKNLDKANLEELRLRQEADDAARDNDKLVKDNRKLKADLDEA
          RIQLEEEQDAKSHADSSRRLLAEIEELKKKRVAKETSDKQKAQDQKANYQRENESLKA
          DRDSIERRNRDAEROVDLRAQLDDALSRLEKSRAKEKSVEANRELKKVLDRERQS
          LESLSKPNSALESDKQILEDEIGDLHEKNKQLQAKIAQLQDEIDGTPSSRGGSTRGAS
          ARGASVRAGSARAEE"
     intron        896..1160
                   /note="intron I"
     intron        1311..1380
                   /note="intron II"
     intron        1566..1742
                   /note="intron III"
     misc_feature  5799..5804
                   /note="pot. polyA signal"
ORIGIN

Alignment Scores:
Pred. No.:            0.0587           Length:           5894
Score:               180.00           Matches:           132
Percent Similarity:  39.02%           Conservative:      131
Best Local Similarity: 19.58%         Mismatches:        245
Query Match:         5.47%            Indels:            167
                                      DB:                 29
                                      Gaps:                3
```

US-10-736-868-2 (1-643) x ACMHC (1-5894)

```
REFERENCE
AUTHORS    Jones,L., Murphy,L., McNeil,A., Simpson,I., Harris,D.,
           Barrel,B.G., Rajandream,M.A. and Walsh,S.V.
TITLE      Direct Submission
JOURNAL    Submitted (23-OCT-1995) Schizosaccharomyces pombe chromosome I
           sequencing project, Sanger Centre, Hinxton Hall, Hinxton, Cambridge
           CB10 1RQ E-mail: barrell@sanger.ac.uk
COMMENT    Notes:
           Details of yeast sequencing at the Sanger Centre are available on
           the World Wide Web.
           (URL, http://www.sanger.ac.uk/Projects/S.pombe/)
           Protein coding regions (CDS) have been predicted with the help of
           computer analysis using the Genefinder program in PomBase (an ACEDB
           database) with additional predictions for the branch-acceptor sites
           supplied by the program Sp3splice. CAUTION: It is possible that for
           any individual CDS we may have underestimated or overestimated the
           number of introns/exons or we may not have chosen the correct
           splice donor/acceptor sites. CDS are numbered using the following
           system eg SPACSH10.01c. SP (S. pombe), A (chromosome 1), c5H10
           (cosmid name), 01 (first CDS), c (complementary strand).
           The more significant matches with motifs in the PROSITE database
           are also included but some of these may be fortuitous. The length
           in codons is given for each CDS.
           IMPORTANT: This sequence MAY NOT be the entire insert of the
           sequenced clone.  It may be shorter because we only sequence
           overlapping sections once, or longer, because we arrange for a
           small overlap between neighbouring submissions.
FEATURES             Location/Qualifiers
     source          1..19514
                     /organism="Schizosaccharomyces pombe"
                     /mol type="genomic DNA"
                     /strain="972h-"
                     /db_xref="taxon:4896"
                     /chromosome="I"
                     /map="IR"
                     /clone="cosmid c29E6"
     gene            1..804
                     /gene="SPAC29E6.01"
     CDS             1..804
                     /gene="SPAC29E6.01"
                     /note="SPAC29E6.01, len:267, SIMILARITY:Caenorhabditis
                     elegans., YSS1_CAEEL, hypothetical 80.3 kd trp-asp repeats
                     containing protein k10b2.1 inchromosome ii., (701 aa),
                     fasta scores: opt: 686, E():0, (38.7% identity in 269 aa)"
                     /codon_start=1
                     /label=SPAC29E6.01
                     /product="hypothetical trp-asp repeat-containing protein"
                     /protein_id="CAA91423.1"
                     /db_xref="GI:1044927"
                     /db_xref="UniProt/Swiss-Prot:Q09855"
                     /translation="DRTVSVWDVNSRFILYKLYGHSGSVLCLDFCRRRNLLVSGSSDS
                     TIIIWDWQNRRPLKVYFGHTDNVLGVVVSENYIISSRDHTARVWRLDATSPAEACMH
                     VLRGHLASVNSVQYSSKTGLIVTASSDRTLRTWDITTGHCIRIIHAHQRGIACAQYNG
                     KFIVSGSSDLITIRIFEASSGKLLRMLQGHEDLIRTVRFNDEKIVSGGYDGTVRIWNFN
                     TGEQHCVLHNSRNSRVFGLQFDHRRIIACTHSSEILVWNFDDGLDCTFF"
     misc_feature    40. .150
                     /gene="SPAC29E6.01"
                     /note="Match to PF00400 WD40, WD domain, G-beta repeat
                     Score 42.94"
     misc_feature    163. .267
                     /gene="SPAC29E6.01"
                     /note="Match to PF00400 WD40, WD domain, G-beta repeat
                     Score 34.79"
     misc_feature    298. .408
                     /gene="SPAC29E6.01"
                     /note="Match to PF00400 WD40, WD domain, G-beta repeat
                     Score 42.65"
     misc_feature    418. .525
                     /gene="SPAC29E6.01"
                     /note="Match to PF00400 WD40, WD domain, G-beta repeat
                     Score 20.89"
     misc_feature    538. .648
                     /gene="SPAC29E6.01"
```

```
Qy    342 GlnLysSerGlnLeuValLeuLeuGlyPheGlyLeuHisGlyPheGluSerAspAspAspGluAsp 361
             ||||||                                     ||:::::|||
Db   4471 -----------CAGCTC------------------------------GGCCAGCAGCTCGAGGACGGCCGC 4500

Qy    362 GluGluAspGluAsnLeu---------IleAspProSerGluAsnSerPheArgArgAla 378
             :::|||                   |||  ||||||
Db   4501 TCCGAGGTCGACTCGCTCAAGAGACAAGCTCTCGGCCCCCGAGAAGTCGCTCAAGACCGCC 4560

Qy    379 ProLeuValLeuSerSerGlyPheValGluLysLeuSerLeuLysAsnAspGluLeuLysSer 398
             :::      ||||||:|||:|||:
Db   4561 AAGGACCAG--AACCGCCGACCTCGACGAGCAGCTTGAGGACGAGGCACCGTGCCGCGCC 4617

Qy    399 AlaLeuAspArg---------IleLeuTyrArgValAspAspValGluLysTyrLeu 414
             :::|||:   :::::  :::::|||
Db   4618 AACGTCGACAAGCAGAAGAAGCGCGTCGAAGCTCGAGGCCAAGCTCACCGAGGACCAGGTC 4677

Qy    415 Ala-ProLysProMetGluPheAsnProLysProGlyTyrPheAlaProArgGly 434
             |||                |||
Db   4678 ACCGCCCTCGACGCGCCAGAAGAACGCGCCGCCG---------------CCCAGGCC 4719

Qy    434 sIleProThrArgProArgLysMetLeuProLeuLeuIleGlySerAspProLysValGl 454
             :::||||||||||||||               |||
Db   4720 AAGACCCTCAAGACCCAGGTCGACAGAGACCAAGCGCCGCCTGGAGGAGCCGAGGCCTCG 4779

Qy    454 nGluGluIleArgArgHis---------ProSerThrGluTrpLysIleAlaAlaLysGl 470
             :|||                |||
Db   4780 GCGCCGCGCCGCTCGGAGGAGCAGCGAGGCCGCAGAACGCCCTCGACGAGGTGG----- 4825

Qy    470 uSerArgValLeuThrAsnLeuLysAsnAsnProSerLeuAlaAlaLeuPheMetAlaAspAs 490
             |||                                       |||
Db   4826 -----------------CCCAG-CTCACCGCC--GACCTCGACGC 4850

Qy    490 pLysLeuGluAsnThrLeuLysGlyLysArgGlnMetLeuThrLeuThrAspGluGlnLysGlyArgTh 510
             :::   :::::|||
Db   4851 CGAGGAGGTGAAGCGCCTCGAGGGCGAGCTC--AACAC 4892

Qy    510 rArgValLysThrIleArgAla---------LeuProArgLeuPheGlyAlaProTh 526
             |||||:   ::::::|||
Db   4893 CCGGATCAGGAGCTCCAGTCGAGACGCCGGAGAACGCCGGCCGGCCGCCGTCGTC 4952

Qy    526 rAlaLysAlaGluMetIleAspAlaLysValPheGlnAspIleGluGlnLysArgProIle-P 546
             :::  ::: |||   ||||||||||
Db   4953 CGAGGAGGTGAAGCGCCTCGAGGGCGAGCTC--GAGCGCCTCGAGGAGGAGCTCCTCAC 5009

Qy    546 roProLeuPhePheGluProLysGlyArgHisThrArgTrpThrGlyAlaAsnG 566
             |||||                  |||||||::::::
Db   5010 CGCCCAGGAGGCCGCGCCGCCCGAGAAGAACCTTCAGCCTCGA 5069

Qy    566 luLysGluIleProGlyLeuGlySerArgPheIleLeuProSerLeuAspProThrMetP 586
             :::|||               |||
Db   5070 TGCCGCAGGAGGCGCCGACCACGC 5096

Qy    586 roAlaLeuAsnThrAlaPheSerThrGlnGlyArgAla 598
             |||
Db   5097 CGCGCCGCGACAACGACAAGCTCGTCAAGGACCAACCGCA 5134
```

```
RESULT 8
SPAC29E6/c
LOCUS       SPAC29E6                19514 bp    DNA     linear   PLN 18-OCT-1999
DEFINITION  S.pombe chromosome I cosmid c29E6.
ACCESSION   Z66525
VERSION     Z66525.1  GI:1044926
KEYWORDS    coiled-coil; cysteinyl-tRNA synthetase; G-beta repeat; p115;
            peptide methionine sufoxide reductase; tdf1; TFIID; transcription
            initiation factor Iid; trp-asp repeat; U4/U6 small nuclear
            ribonucleoprotein; U4/U6 snRNP; vesicular transport factor; WD
            domain.
SOURCE      Schizosaccharomyces pombe (fission yeast)
ORGANISM    Schizosaccharomyces pombe
            Eukaryota; Fungi; Ascomycota; Schizosaccharomycota;
            Schizosaccharomycetes; Schizosaccharomycetales; Schizosaccharomycetaceae;
            Schizosaccharomyces.
```

misc_feature
```
                    /note="Match to PF00400 WD40, WD domain, G-beta repeat
                    Score 48.28"
607..651
                    /gene="SPAC29E6.01"
```

gene
```
                    /note="PS00678 Beta-transducin family Trp-Asp repeats
                    signa ture"
join(1228..1308,1359..1517,1564..2223,2266..2576,
2660..3077)
                    /gene="SPAC29E6.02"
```

CDS
```
join(1228..1308,1359..1517,1564..2223,2266..2576,
2660..3077)
                    /gene="SPAC29E6.02"
                    /note="SPAC29E6.02, len:542, SIMILARITY:Homo sapiens,
                    O43446, u4/u6 small nuclear ribonucleoprotein hprp3, (683
                    aa), fasta scores: opt: 946, E():0, (35.2% identity in 542
                    aa)."
                    /codon_start=1
                    /label=SPAC29E6.02
                    /product="putative u4/u6 small nuclear ribonucleoprotein"
                    /protein_id="CAA91424.1"
                    /db_xref="GI:1044928"
                    /db_xref="UniProt/Swiss-Prot:Q09856"
                    /translation="MDRNKRRLESSQLRESSSPNSQNKPNAMEEIKRRRLQLEQRLAQ
                    QAQVPWEKRSENGNNAGMETIQNRISELKERTKRFNANIPKSEGLFRDDSNGAKGGL
                    KVGIHPVLLDGNIQNTILTPENRKRTASFSTKGVSLSQHQLLKPPAITEQNPFLDTAP
                    TPRLEDSPFYDPRTKESRKTRGSRNLHLNESGKFIEEANQARRQARLEDLKKRIALHS
                    HKAGIEDELDITSKSIGRDTIPIEWWDLPFIKDYNDYGDENNWLIDGPQSIINSAIQ
                    HPIPVLPPYAKNQPSHSVFLTKKEQKIRRQTRAEARKEKQDRQLLGIEPPEPPKVK
                    LSNLMHVLGDDAIKDPTKIEAEVRKQVEERRLRHERENEERKLTPBERKEKAFRKKDE
                    DSAAGLRCLVFRIKYLAHRPHRLKIDLNAKQWGATGVCILNANFNIVIFRAGQKAIKK
                    LKRLMLERIPWTDTSRNSIIAQGNKLVDTEGRELNYTENTCNLVWEGEIGRRAFRYWS
                    FRSCPSENDAKSYLEBQGGAEHFWMLAKSWSENV"
```

misc_feature
```
1309..1314
                    /gene="SPAC29E6.02"
                    /note="gtatgt, splice donor sequence"
```

misc_feature
```
1345..1358
                    /gene="SPAC29E6.02"
                    /note="ctaattcaatatag, splice branch and acceptor"
```

misc_feature
```
1518..1523
                    /gene="SPAC29E6.02"
                    /note="gtaagt, splice donor sequence"
```

misc_feature
```
1550..1563
                    /gene="SPAC29E6.02"
                    /note="ctaacgtttgtag, splice branch and acceptor"
```

misc_feature
```
2224..2229
                    /gene="SPAC29E6.02"
                    /note="gtaaat, splice donor sequence"
```

misc_feature
```
2251..2265
                    /gene="SPAC29E6.02"
                    /note="ctgattattttttag, splice branch and acceptor"
```

misc_feature
```
2393..2397
                    /gene="SPAC29E6.02"
                    /note="sequence duplicated by E. coli transposon tn1000,
CA UTION: tn1000 (gammadelta) transposon sequence and
duplicat ed 5bp sequence deleted from this sequence"
```

misc_feature
```
2577..2582
                    /gene="SPAC29E6.02"
                    /note="gtatgt, splice donor sequence"
```

misc_feature
```
2647..2659
                    /gene="SPAC29E6.02"
                    /note="ctaacttattag, splice branch and acceptor"
```

gene
```
complement(join(3131..4432,4588..6237,6312..6479))
                    /gene="SPAC29E6.03c"
```

CDS
```
complement(join(3131..4432,4588..6237,6312..6479))
                    /gene="SPAC29E6.03c"
                    /note="SPAC29E6.03c, len:1039, SIMILARITY:Bos taurus,
P115_BOVIN, general vesicular transport factor p115, (961
aa), fasta scores: opt: 608, E():1.2e-24, (23.3% identity
in 980 aa)."
                    /codon_start=1
                    /label=SPAC29E6.03c
                    /product="putative vesicular transport factor"
                    /protein_id="CAA91425.1"
```

```
                    /db_xref="GI:1044929"
                    /db_xref="GOA:Q09857"
                    /db_xref="UniProt/Swiss-Prot:Q09857"
                    /translation="MDIFHKSYNVILSPPKVQQADETISKLCDRLEHATLFEDRKAAA
                    LGIKSFAREFKELVAANGLKGIIQSLHRDYDDPELLKVTLETCLILTRHDDDSRASDT
                    GLWIADQFILNQDNIQCLLQLSISHKDFVVRLYSVELFSAILSCRPTELKDCLQTFPSA
                    ISSIMVPLRDSIEPVRNADLYFLSELVKDCTSIQKLVVPENAFETLFSLLENEGGVDG
                    GIIALEALKFLNVLLKDNISNQNYFRESNHIPSLLKLLSVDTFVDGTWDTSRVQCVIE
                    ALYALQSLVPIGLSSSTANQNAVVSNHGIDVLLTLATHPDLLFFDVQKISWITLAHVV
                    YSNARAQNSFVDSTFFDIKNTDVLTCLFDYLFLSSISPSHRYSVAFFLRSLTSENDEL
                    SSKFLKQIIHAYTHKQDNRLNIIQGYLDLVHLSDQDYDNWFTSTILYTLYIVIDNDQRK
                    YLLCSIPLFQDMDNDEDSESEDKVTFIQCVSTKLIATLRHENALQNCVGYLTLLIALV
                    YGNPDSVKDFLSESSILQTFLTALMDESSSANSVIQGMIAVFLSLVYYYCPIESPVSK
                    SDVYNATSAVKRDVFINRLQRLRRMNLYEITFLSMKQKMQLKSLREEIDNTKEALDL
                    SVKERSIOEEKLNESLKTSKTNLEEQTQLAEKYHEELLDNQOKLYDLRIELDYTKSNC
                    KQMEEMQVLREGHESEIKDFIEEHSKLTKQLDDIKNQFGIISSKNRDLLSELEKSKS
                    LNNSLAALESKNKKLENDLNLLTEKLNKKNADTESFKNTIREAELSKKALNDNLGNKE
                    NIISDLKNKLSEESTRLQELQSQLNQDNQIETLNERISAAADELSMESINKNQANE
                    LKLAKQKCSNLQEKINFGNKLAKEHTEKISSLEKDLEAATKTASTLSKELKTVKSEND
                    SLKSVSNDDQNKEKSVNNEKFKEVSQALAEANEKLNARDEEIERLKVDIIGLQNASLN
                    MQSLKDSDNRTISDLESKNKELEKKLKEADEVWLLIVEBELESKRTKDKELLRQCGQAV
                    SEDEQSEBE"
```

misc_feature
```
complement(4433..4449)
                    /gene="SPAC29E6.03c"
                    /note="ttgacgaagcacaaag, splice branch and acceptor"
```

misc_feature
```
complement(4582..4587)
                    /gene="SPAC29E6.03c"
                    /note="gtaatt, splice donor sequence"
```

misc_feature
```
complement(6238..6250)
                    /gene="SPAC29E6.03c"
                    /note="ctaacaaatttag, splice branch and acceptor"
```

misc_feature
```
complement(6306..6311)
                    /gene="SPAC29E6.03c"
                    /note="gtatga, splice donor sequence"
```

gene
```
6914..7351
                    /gene="SPAC29E6.04"
```

CDS
```
6914..7351
                    /gene="SPAC29E6.04"
                    /note="SPAC29E6.04, len:145, SIMILARITY:Saccharomyces
cerevisiae, NNF1_YEAST, nnfl protein, (201 aa), fasta
```

Alignment Scores:
```
Pred. No.:                0.305       Length:              19514
Score:                    179.50      Matches:             131
Percent Similarity:       37.34%      Conservative:        124
Best Local Similarity:    19.18%      Mismatches:          289
Query Match:              5.45%       Indels:              139
DB:                                   Gaps:                28
```

```
us-10-736-868-2 (1-643) x SPAC29E6 (1-19514)

Qy      17  ProLeuSerAlaGlnSerProSerThrSerAspAlaProGlyAlaLeuSerSerLeu  36
                ||||||  |||||||  |      ||||||            ||  |||
Db   18056  CCTTTTCAGCCTCCTTATCCCTCTAAAGCAGATGAGAAAAATAGCTATCATTCTGATCTT  17997

Qy      37  ValGlyLysSerHisGlnLysLeuPro----------LeuAlaProSerMetGluAla  52
                :::::::::||||||||                  ::::||:||
Db   17996  TATAACTTCGGGACTAGTTTAACGGTGAAAGGAGTATCTTAACGGTAGCGGATGACCTA  17937

Qy      53  LeuGluLeuMetGlyValGlnPheValAspAlaLeuIleLysGlyGlnMetGluMet  72
                |||:::  |:::  |||  :::|||||  ||||||  |:::
Db   17936  TTGAAGATGATGGGAACAGTTTATAGAAATGATGGAACAGTTAGCTGAGCGTCGTATG  17877

Qy      73  AlaLys------------------------------------GlyAlaPheLysThrGlnLeuGlyVal  83
                :::                                  |||||||
Db   17876  CAGCGTGAAGATAATAGCAATTTCACGAACCAGAGTTATACAGAAAGCCGGTTGAATAT  17817

Qy      84  LeuGluLysValHisProAspGlnPheAspLysThrTyrLysLysLeuLysValAsp------  101
                ||||::  ::::|||:::|||  :::|:::
Db   17816  GACCAAGATGAAGAAGACGAAGAAGAT--------------------GTAGATGAGGAC  17775

Qy     102  ------AspLeuAlaAlaAspAlaValMetGlnGlnAlaGlnMetAlaGluGlnLeuGlnPro  119
                     ||:::                            :::
```

```
Db  17774 GAGTTAGATCTTAATGACCGATGAACAAGATGGAAGAAGACGCAGGATGTTCCAAATT 17715
Qy  120 LysSerGlyAsnAlaPheIleAspMetLeuLeuAsnGlyAsnGlyIleProIleGlySerSer 139
Db  17714 TTTGCTGCCCGTTTATTTGAGCAACGTGTG------- 17685
Qy  140 IleArgGlyLeuGluGluArgThrGlnArg 151
Db  17684 TTGCAAGCTTATCGTGAGAAAGTTGCTCAACACGGCAAGTCAAGCTAAACTTTTAGAAGAGATT 17625
Qy  152 AspMetGluAsnThrArgProSerGluProGlnIleIleAlaLysAlaValMetAspLysPheGln 171
Db  17624 GAGGAGGAGACAAACCAAACAGGACGTGAGCTGAGCTCAAAAAAAATAAGAGAGAAGGAGAAA 17565
Qy  172 ThrGlnIleLeuProGlyLeuValAlaAsnMetIleAlaGlyLysLysAsnProPheLysMet 191
Db  17564 AAGAGG-----------GATAAAAAGAGAGCAATTGAAATTG 17535
Qy  192 ProGlnGlnMetArgLysAlaGlnAlaAlaAlaProSerValPheGlnGlnAlaLeuAla 211
Db  17534 GCGAAGGAGGAAGAGGCGTCAAAGACACGGAAGCTCTTGCTGAACAGGCT---GCT 17478
Qy  212 GlnArgAlaMetLeuGlyLysAsnAlaProValAlaGlyValGlyGlyGluGlnArg 231
Db  17477 CAAAAGCATTGGAAGCTAAAGACAGAGAAGACAGAGAAGACTCGCGACAACGGGAGAGCAACGT 17418
Qy  232 MetMetMetAsnArgValAspGlnInArgMetGlnGln----------ArgGluLeuGluGlu 248
Db  17417 ------------TTGAAGCGGTGAGCAGGAGAAAAGCAGCAACTTGAGCCTCAGAAAAGAGAG 17364
Qy  249 GluAspGluAspAspAspAspGluLeuLeuGluAspValProArgArgArgSerSerAsp 268
Db  17363 GAAAAGCAAAAGCAGAAGGAAAAGAGGAAGAGAAGAAACTCAGAAGCAGCAACAGAGGCTGAC 17304
Qy  269 GlyGluGlnSerGluAlaGluHisGlnAlaGlnArgArgAspArgGluAlaArgArgLeuLysSer 288
Db  17303 CGTGAAAAAATGGCAGACAACGACTAAGAGAGAAGAAGAGAGAAAAACGTATTTTGGAA 17244
Qy  289 SerProArgLeuLysLysGlnLeuLeuGlnAsnAlaGluValGlnAsnSerLeuSerTyrGln 308
Db  17243 GAAAAGGAAGGCGAAGGGAAAAATTAGACAAAGAAGAAGAGAA 17202
Qy  309 ArgMetArgAspSerProLeuSerLysArgProLeuAlaMetAsnAspGlu 328
Db  17201 ---------CGTCGGCGTCGTGAGTTACTTGAAAAAGAATCAGAAGAG 17163
Qy  329 SerAlaPheArgAlaMetGluAlaArg--AlaLysLeuAspGlnLysSerGlnLeuVal 347
Db  17162 AAAGAAGAAGGAGATTAAGAAGCTAAGATCGCTGCTTTTCTTTGCACCTAAATCAGACTAAA 17103
Qy  348 LeuGlyLeuHisGlyPhe-----------------GlyGlu 355
Db  17102 GAGGGGTTCTGATGGTTGCTACTTCTTCTCAACTTGGCCTTTTGAAAAAAAGGCGAT 17043
Qy  356 SerAspAspGluGluAspGluGluAspProSerGluSerPhe 375
Db  17042 CTCGTAAACGATGAGGATAAGCTTTCTTCACATCTCCTAGATTCTGTACCTAATGTGTTTG 16983
Qy  376 ArgArgAlaProLeuArgLeu-----SerSerGlyPheValGlyLeuLysLeuLysSerAsn 393
Db  16982 CGGCAAGCTCCCATAGGCCTTAAGAACACAAATAATTTATCGGAGGCGCAATGCTCTTCTCT 16923
Qy  394 AspGluLeuLysSerAlaLeuAspArgIleLeuTyrArgValAspAspValGluLysTyr 413
Db  16922 AATCTGCTTAATTCTTCGTTATTTTCATCTTTAATTCCGTCACCCTTTAATTTCT-- 16866
Qy  414 LeuAlaProLysProMet-----------GluPheAsn 422
Db  16865 CTTGAGCTAATCCTTTAAATAGAACACTTAATAATTCCGTAAACTTAACTGATTTTGGT 16806
Qy  423 ProLysProGlnProGlyTyrPheAlaProArgLysIleProThrArgProArgLysMet 442
Db  16805 AGAAAGCCAAAT---GGTTTACATTCTCCATCTCACTCCTCCTAAGTAATTCCAACAAT--- 16752


Qy  443 LeuProLeuLeuIleGlyGlySerAspProLysValGlnGluGluIleArgArg-----His 460
Db  16751 -------------TTTGATTAATCCAAATGCTGGCATTCCTCCTAGTAGAGCCAATTCA 16704
Qy  461 ProSerThrGluTrpLysIleAla---------------------LysGluSerArgValLeu 474
Db  16703 CCGGTGCATCATTATCCTTTGCAACACCTCCTTCCAACGTGCTAATAAATATCCATTG 16644
Qy  475 ThrAsnLeuLysAsnAsnProSerLeuAlaAlaLeuPheMetAspAspLysLeuGluAsn 494
Db  16643 AATAATGGAGCAAACGTACCAGCTTTGCTTAATTCATTTCAAGTCCACAATTAAGTCCT 16584
Qy  495 ThrLysLysGlyArgGlnMetLeuThrAspGluGlnLysGlyArgThrArgValLysThr 514
Db  16583 CTAGTA------AATAGAGTTTGAACGAGCCGTCTAGTTCCCCTTTATCGTCCTCAAGT 16530
Qy  515 IleArgAlaLeuProArgLeuPheGlyAlaThrAlaLysGluMetIleAspAla 534
Db  16529 TTGAAA-------TCCCCGTTATCCAAGGAAGGGTCTTAAATCAA 16491
Qy  535 LysValPheGlnAspIleGlu--------GluArgProIleProPro 547
Db  16490 CAAGGACATGAGCAGTATAACTTCTCACTGTCACCATCTATCCGAAATAAGCTATCTCCT 16431
Qy  548 LeuPhePheGluProLysGlyLeuGlyArgHisThrArgLeuArgTrpThrGlyAlaAsnGluLys 567
Db  16430 ATTTGTCGTCCTCCAGTCCAAGGAGGTTCTCCTAAGCTGAAGAATAATTTGTCCAACACAGAG 16371
Qy  568 GluIleProGlyLeuGlySerArgPheIleLeu---ProSerLeuAspProThrMetPro 586
Db  16370 GAA------AGGATGGGTAGTCGTGCACTATTAGATGACAAAACAGATTCTGTTATCACA 16317
Qy  587 AlaLeuAsnThrAlaPheSerThrGlnGlyIleArgAlaArgAspArgTrpAspThrMetPhe 606
Db  16316 GCTTCTAATTCA------ACCACTTCTGGTCTGTCACGAGATGAATCTAGTAAT------ 16269
Qy  607 LysIleProAsnAsnTrp----------AsnProGlyAspGluValGlyPheLysMetAsn 623
Db  16268 ------CCTAAATAATTATGAGTTGTTAAACGCGTTTAATCAAACACCTGGAAAAATATCT 16215
Qy  624 SerLysThr 626
Db  16214 CGGTCGACT 16206
```

Fuchs,M., Dusterhoft,A., Fritzc,C., Holzer,E., Moestl,D.,
Hilbert,H., Borzym,K., Langer,I., Beck,A., Lehrach,H.,
Reinhardt,R., Pohl,T.M., Eger,P., Zimmermann,W., Wedler,H.,
Wambutt,R., Purnelle,B., Goffeau,A., Cadieu,E., Dreano,S.,
Gloux,S., Lelaure,V., Mottier,S., Galibert,F., Aves,S.J., Xiang,Z.,
Hunt,C., Moore,K., Hurst,S.M., Lucas,M., Rochet,M., Gaillardin,C.,
Tallada,V.A., Garzon,A., Thode,G., Daga,R.R., Cruzado,L.,
Jimenez,J., Sanchez,M., del Rey,F., Benito,J., Dominguez,A.,
Revuelta,J.L., Moreno,S., Armstrong,J., Forsburg,S.L., Cerutti,L.,
Lowe,T., McCombie,W.R., Paulsen,I., Potashkin,J., Shpakovski,G.V.,
Ussery,D., Barrell,B.G. and Nurse,P.

TITLE   The genome sequence of Schizosaccharomyces pombe
JOURNAL   Nature 415 (6874), 871-880 (2002)
MEDLINE   21848401
PUBMED   11859360
REFERENCE   2 (bases 1 to 36315)
AUTHORS   McDougall,R.C., Rajandream,M.A., Barrell,B.G., Brown,S., Murphy,L.,
Jones,L., McNeil,A. and Harris,D.
TITLE   Direct Submission
JOURNAL   Submitted (17-JAN-2000) European Schizosaccharomyces genome
sequencing project, Sanger Institute, The Wellcome Trust Genome
Campus, Hinxton, Cambridge CB10 1SA, E-mail: pombe@sanger.ac.uk
COMMENT   Notes:
Details of S. pombe sequencing at the Sanger Institute are
available on the World Wide Web.
(URL, http://www.genedb.org/genedb/pombe/index.jsp) (URL,
http://www.sanger.ac.uk/Projects/S_pombe/)
CDS are numbered using the following system eg SPAC5H10.01c.SP (S.
pombe), A (chromosome 1), c5H10 (cosmid name), .01 (first CDS), c
(complementary strand). However, clones may have been reorientated
since the original submission, therefore the complementary strand
notation MAY be invalid for strand inference. IMPORTANT: This
sequence MAY NOT be the entire insert of the sequenced clone. It
may be shorter because we are only sequence overlapping sections once.

FEATURES   Location/Qualifiers
source   1..36315
    /organism="Schizosaccharomyces pombe"
    /mol_type="genomic DNA"
    /strain="972h-"
    /db_xref="taxon:4896"
    /chromosome="I"
    /map="IR"
gene   /clone="cosmid c30"
    complement(253..5721)
    /gene="SPAC30.01c"
CDS   /note="synonyms: sec702, sec7b"
    complement(253..5721)
    /gene="SPAC30.01c"
    /note="guanine nucleotide exchange factor (predicted);
    involved in the secretory pathway (predicted); Sec7
    domain; similar to S. cerevisiae SEC7; non-clathrin
    vesicle coat component (putative); involved in
    intracellular protein transport"
    /codon_start=1
    /protein_id="CAB66460.1"
    /db_xref="GI:6723884"
    /db_xref="UniProt/TrEMBL:Q9P7V5"
    /translation="MQDASEPVINMPALPMNSVNEQHDRHEDLPEVGTTSVTKIINDS
    DNEBDDENGMDLGRVASESLEGDAVVSIDINTEDSSLSPAKQENEKSPEGIEQKYQEED
    LKDDKKSNETIATPVDPAASRASLEKQKSGLTNLEKSTHVVLIKCIEQLIQIKVCMKN
    EKMKNLMEQLKPLLQTECQFDKFLIVELFQYCFESSQDEVMNISLDTISKLASFAYFS
    SKDKTPASFGPPKSLLQCMVDMVCDSINDEVVDGNLQLNVKALSAFIICSEQDSMLH
    GAILLNSVRKLFLNVFLLGDSDTIQSVAQASLTQAVTVVERLRASHTQSNSTSALPEE
    DASVTENWVHDEDEPDKKITIHSMASAGTSSLDHVKVDADDPAVTSVENSSIQDAFLV
    FRSMCRLAVRQTSPDKVSNIRSQAMRAKLISLHLYRILEKNSDLFMDPTLQFRGIPA
    LKGMTLVHASRQYICLVLSRNAVSPVPVEVCCDIFYLMVFSLRAHFKQEIEVFFRE
    VYFPMLDLKNTSYNQKLHTLLIIQRICLNPRALVELYINVDCDRSSTTNVFEQLLFSI
    SKVTTNGPSETISEDIEEILPSLESSERSSTPFLNTNSASLKSEVVQLTTFSDFQLKL
    KTLQCVLDIIQSLSNWAESGLYLSRRGVSTDEQGFVGDYDALSRSDTPVTNPYYNGKQ
    SFEANSHSSSSIALADPSQFESNKQRKLLRTCINKFNYKPTRGLKMLSENEYVDIND
    PKAIAEFLFRADGIDKTTLGDYLGEGDEKSISVMHEFIDCLSFINLKFVDALRRLLQC
    FRLPGEAQKIDRIMLKFSERYMKENPSAFANADTAYILAYSIILLNTDLHSPRIKNQM
    TKEDFIKNNRGINDGADLDEDYLGFVYDDILKNEIAMKDDQELAAIAPLMNNFSTSSG

FTTFTSNGRDLQRVACIQASEMANKATSVLKLLYQQKHGSOXTNVYNATHFEHIG
PMLEATWMPILAALSNPLQNSDYVNELNMCLDGFQLVRIACLFDLLIRDAFIKTLT
NFTNLHSTSEIKLRNTMVIKTLRIASTEQNVLKIASTEQNVVDSVQKDILTIISQLERVQLIGVGVD
ETEVPDVINARVRRKNVNIGSSNSINRHVSGSTSRSTRTRSLSKPLSPEAVSELMSTEV
VLSIDRIFTQTSSLSGSAIVSFFKALCEVSWDEITSSSDLEQPRLYSLQKLVRISYYN
MQRIRVEWSSINVLGRFPNMVGSDENRHVAVFALDSLRQLSMHFLEIEELSLFSFQK
EFLKPFEVMASDTVVEVKELVLQCVKQMIQAKISKIKSGWKTLFGVFTPAAKARSEI
LISMTFDTLVNLFSEHYDTLMQQNCLIDMLISFTELCKNGTNQKISLQSLEIIREVVS
SLSTMIKEGLSSKPSVNETPSKYVFPVLFAYDIIMSAEDLEVRSRALQNLFIFLEE
SDDFTESTWEVVSRKFIFPIFSIFGPEADEATVMLRDEIRTWQSTTIVEALRSLVTL
LTRRFDKLHNLLKGYLWLFSNCICRDNITLSRIGTNCMQQLLSGNAYRFEVKDWNLVA
DMFIELFKETTPHQLLLLETFSNGOGAPVYSENENTQLSHKRGGSLPETSRSISTSSI
SPEKQMEPRSMIRKCIIQLLJLISIVABLLDNEEVFNHIPHEHVLKITTVAIYDSWQFAR
KFNEDKSLRITLLNVGFMKQLPNLLRQETASALLYITLLFRLLKTRDPLGKTETDQKI
HKLLFPVCAEMLDMYASLVEKHTRNHAAWQPVIATILDSIINLPLELFSENIHTLYF
SCCSMIAKENLDDQLRELLKNYFNRVGHILLNLNAQQE"

misc_feature   complement(3049..3609)
    /gene="SPAC30.01c"
    /note="Match to PF01369 Sec7, Sec7 domain Score 329.12"
3'UTR   complement(610..6195)
    /note="mRNA from AU011189"
gene   complement(join(6196..6869,7015..7186))
    /gene="SPAC30.02c"
CDS   complement(join(6196..6869,7015..7186))
    /gene="SPAC30.02c"
    /EC_number="2.7.1.78"
    /note="polynucleotide kinase (predicted); RNA polymerase
    II elongator associated protein (predicted); similar to S.
    cerevisiae KTI12"
    /codon_start=1
    /protein_id="CAB66461.1"
    /db_xref="GI:6723885"
    /db_xref="UniProt/TrEMBL:Q9P7V4"
    /translation="MPIIVSGYPSSGKTTRSNELKKALEDRIHQNIDNTKDYRVIII
    NDESLNIEKETYRESKNEKAARGLLYSAVORELSKSTFVICDALNYIKGFRYQLYCES
    KSMYTHCVIHVAVPQDLCRKFNSNKEQPYPDDVLEQLMFRYEEPNGMTRWDSPLFTV
    LHDDASCPIDDIWSVLIHNKNVKPNQATMVKPPAEVNYLYELDKTTQDVIMLILDNSN
    DTSLIVPGSKLQIALPSVTVSLPLLQRLRRQFIQINRQQSYNTNVLKEMFVEFLNGQ
    FETLD"
misc_feature   complement(6870..6883)
    /gene="SPAC30.02c"
    /note="ctaattgtcttag, splice branch and acceptor"
misc_feature   complement(7009..7014)
    /gene="SPAC30.02c"
    /note="gtatgt, splice donor sequence"
gene   complement(join(8004..8038,8096..8264,8307..8389,
    8428..8788,8827..8889))
    /gene="SPAC30.03c"
CDS   complement(join(8004..8038,8096..8264,8307..8389,
    8428..8788,8827..8889))
    /gene="SPAC30.03c"
    /note="translin-like protein; DNA-binding protein;
    involved in meiotic recombination (predicted);
    non-essential (PMID 12618370); no apparent S. cerevisiae
    ortholog"
    /codon_start=1
    /protein_id="CAB66462.1"
    /db_xref="GI:6723886"
    /db_xref="UniProt/TrEMBL:Q9P7V3"
    /translation="MNKSIFIQLQDQDIKEHSIREKLTAEVDLLDEKLRVLQLLANC
    EQSRNENLQEKEHGLTLEDLENQEEILEALEIIKSKTRGLAEBLASNFPYYKNGWDR
    SIQKVVYLYLLASWTGRLDKSLRPTYSLLSLSEVGQILQVPVFPESTFHLSIEQYLH
    AVLSELARQSVNSVIGNYHIPFEALNTIQKVHSSFQVLSLKNDSLRRHFDGLKY
    DLKRSEDVVYDLRIHKIV"
misc_feature   complement(8039..8051)
    /gene="SPAC30.03c"
    /note="ctgcatatgcag, splice branch and acceptor"
misc_feature   complement(8090..8095)
    /gene="SPAC30.03c"
    /note="gtatga, splice donor sequence"
misc_feature   complement(8265..8280)
    /gene="SPAC30.03c"
    /note="ttgcacaatttcttag, splice branch and acceptor"

```
misc_feature   complement(8301..8306)
               /gene="SPAC30.03c"
misc_feature   "gtatat, splice donor sequence"
               complement(8390..8403)
               /gene="SPAC30.03c"
misc_feature   "ccacaatcttcag, splice branch and acceptor"
               complement(8422..8427)
               /gene="SPAC30.03c"
misc_feature   "ggcgt, splice donor sequence"
               complement(8789..8800)
               /gene="SPAC30.03c"
misc_feature   "ctaacattacag, splice branch and acceptor"
               complement(8821..8826)
               /gene="SPAC30.03c"
gene           "gacta, splice donor sequence"
               complement(9229..13638)
               /gene="SPAC30.04c"
CDS            complement(9229..13638)
               /gene="SPAC30.04c"
               /note="ABC transporter family; unknown specificity; 15
               predicted transmembrane helices; similar to S. pombe
               SPAC9E9.12C"
               /codon_start=1
               /protein_id="CAB66463.1"
               /db_xref="GI:6723887"
               /db_xref="GOA:Q9P7V2"
               /db_xref="UniProt/TrEMBL:Q9P7V2"
               /translation="MENFHRPFKGGFGVGRVPTSLYYSLSDFSLSAISIFPTHYDQP

Alignment Scores:
Pred. No.:              0.688                    Length:         36315
Score:                  179.50                   Matches:        131
Percent Similarity:     37.34%                   Conservative:   124
Best Local Similarity:  19.18%                   Mismatches:     289
Query Match:            5.45%                     Indels:         139
DB:                                              Gaps:           28

US-10-736-868-2 (1-643) x SPAC30 (1-36315)
```

DB: `e6438bc151f316bc`

```
Qy   495  ThrLeuLysGlyArgMetLeuThrAspGluGlnLysGlyArgThrArgValLysThr 514
          :::
Db 32739  CTAGTA------AATAGAGTTTGAACGAGCCGTCTAGTTCCCCTTTATCGTCCTCAAGT 32686

Qy   515  IleArgAlaLeuProArgPheGlyAlaProThrArgAlaLysAlaGluMetIleAspAla 534
          :::||||:::        ::||||::
Db 32685  TTGAAA------------TCCCCGTTATCCAAGGAAGGCCGTCTTAAATCAA 32647

Qy   535  LysValPheGlnAspIleGlu------------GluArgProIleProPro 547
          :::                           |||
Db 32646  CAAGGACATGAGCAGTATAAACTTCTCACTGTCCACCATCTATCCGAAATAAGCTATCTCCT 32587

Qy   548  LeuPheGluGluProLysGlyArgHisThrArgLeuArgTrpThrGlyAlaAsnGluLys 567
          :::        ::||||::
Db 32586  ATTTGTCGTCGTCCAGTCAAGGAGTTCTCCTAAGCTGAAGAATAATTGTCCAACACAGAG 32527

Qy   568  GluIleProGlyLeuGlySerArgPheLeu--ProSerLeuAspProThrMetPro 586
          ::||||||||||          :::
Db 32526  GAA------AGGATGGGTAGTCCGTGCATTAGTGACAAAACAGATTCTGTTATCACA 32473

Qy   587  AlaLeuAsnThrAlaPheSerThrGlnGlyThrArgAlaArgAspGluTrpAspThrMetPhe 606
          |||:::        ::||||||||
Db 32472  GCTTCTAATTCA------ACCACTTCTGGTCGTCACCGATGAATCTAGTAAT------ 32425

Qy   607  LysIleProAsnAsnTrp----------AsnProGlyAspGluValGlyPheLysMetAsn 623
          |||        :::||||||||
Db 32424  ------CCTAATAATTATGAGTTGTTAAACGCGTTAATCAAAACACCTGGAAAATATCT 32371

Qy   624  SerLysThr 626
          |||
Db 32370  CGGTCGACT 32362
```

RESULT 10
AX151655
LOCUS   AX151655   3543 bp   DNA   linear   PAT 22-JUN-2001
DEFINITION Sequence 260 from Patent WO0138351.
ACCESSION   AX151655
VERSION   AX151655.1   GI:14533609
KEYWORDS
SOURCE   Shrimp white spot syndrome virus
  ORGANISM Shrimp white spot syndrome virus
            Viruses; dsDNA viruses, no RNA stage; Nimaviridae; Whispovirus.
REFERENCE   1
  AUTHORS   Xu,X., Yang,F., He,J., Pham,L.Z., He,M., Ye,Y., Shen,Y. and
            Kodira,C.
  TITLE   Nucleotide sequence of the shrimp white spot syndrome bacilliform
            virus (wsbv), systems containing this sequence and detection kits
  JOURNAL   Patent: WO 0138351-A 260 31-MAY-2001;
            PE Corporation (NY) (US) ; The Third Institute of Oceanography,
            State Oceanic Administration (CN) ; Sinogenomax Co., Ltd. (CN)
FEATURES   Location/Qualifiers
     source   1..3543
                     /organism="Shrimp white spot syndrome virus"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:92652"

Alignment Scores:
Pred. No.:   0.0473   Length: 3543
Score: 177.00   Matches: 99
Percent Similarity: 35.39%   Conservative: 67
Best Local Similarity: 21.11%   Mismatches: 165
Query Match: 5.38%   Indels: 138
DB: Gaps: 20

US-10-736-868-2 (1-643) x AX151655 (1-3543)

```
Qy   176  ProGlyLeuValAlaAsnMetIleAlaGlyLysAsnProPheLysMetProGlnMet 195
          |||||
Db   427  CCTGCGTTATTGGCGCTAGCGGCACAACAAGAAGATCGACAATTATACCTGAAAACCAG 486

Qy   196  ArgLysAla------GlnAlaAlaProSerValPheGlnGlnAlaLeuAlaLeuAlaGln 212
          :::
```

Db   1477 ACAAATCAGTCCTATTTCAACACAACAACAACAATCCATCCTTTTGAGTACAGTGGTT 1536
                   |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   489 AspAspLysGluAsnThrLeuLysGlyArgGlnMetLeuThrAspGluGlnLysGly 508
                   |||||||||| |||||| ||||||||||||
Db   1537 --GGAAAGGTTAACACACATTAGGTGGGAAAGAGGCCGAAGAAAGATTACACAAGACT 1593
                     ---------ArgThrArgValLysThrIleArgArgAlaLeuProArgLeuPhe 522
                   ::::||||||||
Qy   509 -----------
                   ::::||||||||
Db   1594 ATGGAGTCTATAATCTTAAAAACAGGGTAAAAACTCTT----------------- 1632
                   |||||||||
Qy   523 GlyAlaProThrAlaLys---------   -AlaGluMetIleAspAlaLysValPheGln 538
                   ||| || :::||||:::::::
Db   1633 ---CTAGAACAACAAAGAACCTTCAATGCCTGAACTGGTGAAAGTT---GTGTTCCAA 1686
                   ||| |||||| |||||||
Qy   539 AspIleGluGluArgProIleProPro 547
                   ||| |||:::: |||||
Db   1687 GAT---CCAGAAAACCCAGTAAAACCA 1710

RESULT 11
AF369029/c
LOCUS   AF369029   292967 bp   DNA   circular VRL 26-JUL-2001
DEFINITION   White spot syndrome virus, complete genome.
ACCESSION   AF369029
VERSION   AF369029.1 GI:15021392
KEYWORDS   .
SOURCE   Shrimp white spot syndrome virus
  ORGANISM   Shrimp white spot syndrome virus
               Viruses; dsDNA viruses, no RNA stage; Nimaviridae; Whispovirus.
REFERENCE   1 (bases 1 to 292967)
  AUTHORS   van Hulten,M.C., Witteveldt,J., Peters,S., Kloosterboer,N.,
               Tarchini,R., Fiers,M., Sandbrink,H., Lankhorst,R.K. and Vlak,J.M.
  TITLE   The white spot syndrome virus DNA genome sequence
  JOURNAL   Virology 286 (1), 7-22 (2001)
  MEDLINE   21342572
  PUBMED   11448154
REFERENCE   2 (bases 1 to 292967)
  AUTHORS   van Hulten,M.C.W., Witteveldt,J., Peters,S., Kloosterboer,N.,
               Tarchini,R., Fiers,M., Sandbrink,H., Lankhorst,R.K. and Vlak,J.M.
  TITLE   Direct Submission
  JOURNAL   Submitted (15-MAR-2001) Laboratory of Virology, Wageningen
               University, Binnenhaven 11, Wageningen 6709 PD, The Netherlands
FEATURES   Location/Qualifiers
     source   1..292967
                     /organism="Shrimp white spot syndrome virus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:92652"
     gene   1..615
                     /gene="VP28"
     CDS   1..615
                     /gene="VP28"
                     /note="envelope protein"
                     /codon_start=1
                     /product="ORF1, VP28, gene family 1"
                     /protein_id="AAK77670.1"
                     /db_xref="GI:15021393"
                     /translation="MDLSFTLSVVSAILAITAVIAVFIVIFRYHNTVTKTIETHTDNI
                     ETNMDENLRIPVTAEVGSGYFKMTDVSFDSDTLGKIKIRNGKSDAQMKEEDADLVITP
                     VEGRALEVTVGQNLTFEGTFKVWNNTSRKINITGMQMVPKINPSKAFVGSSNTSSFTP
                     VSIDEDEVGTFVCGTTGAPIAATAGGNLFDMVYHVTYSGTETE"
     gene   complement(710..2902)
                     /gene="VP28"
                     /codon_start=1
                     /product="ORF2, putative serine/threonine protein kinase
                     (PK1), gene family 2"
                     /protein_id="AAK77671.1"
                     /db_xref="GI:15021394"
                     /translation="MEGGDQRTKLTPATVMGLYQSKTPGEGEGEGGQFKIPSAIAV
                     KSCCSKNATRRSPPSDSPYSLRPMKRLKKNNGEVGGKAPPVTLRLREDYESTPYNFN
                     RNKKKRTIDENQFATLNPTYATDIIKKQQLPSVSAASVLRKHRANADTQYRKRFSH
                     PNCAKFSTVNLKARDYTPLSVLRSHVKGPKGHLKSSCDTVTETNVVKRNFSSIDKWVKL
                     EKPPCYFAVAEADTNIAAGLESPFHLIRQAAKLGLISDVQDVSSNYETIKQSCIDAKE
                     KASKFLWSNNRTKQPPSSWWPVGFGSKNLSVLDTSPLLNWNRLCKNNGKGWIKTMSID

HMAKNVFKLSPGACESILEKKTLLGEVTAQCKKWESYRRNIPVPAHVQPEYASQVVM
IGPSELYLEVKVGVYYMLETGKVIKFMTDKEMVCEFVFETVFSHALEGRMKGAVGVRK
MCVEBGFCVEMDFAGISVIDVLNGDLKCKMDENVVQQPNPSTTSSKPAAELMQDHGSLC
RMRDTLYGVRMLQATGRLPEGLQSKCKKPITDSISAIAIVGKMRERMLNLQPFVLVEI
VNIVTRLSQQGLVNPDIKSDNIVIDGITGPKMIDFGLIVPCKKYYNFKCWGTDERFF
SNHPHTAPEFINSELCSETAMTFGLAYLLIDMLSSILIKRTADLSANSIYTNIPFLSIV
SKMYDEKTNRPRAYEIAPVIGACPFPFKDNIAKLFQSPKHSLYSKKVK"
complement(3118..4989)
/codon_start=1
/product="ORF3"
/protein_id="AAK77672.1"
/db_xref="GI:15021395"
/translation="MAWTVMALKDAFTERLVVNKVGSGTDMAPVVEDDRQKSLFQKVE
NLYRVLVVEQKNSAITLGNKNTNKRQCRQVEEDKVIFEGEDRTVSNLPQAVKETIAA
NAESILDYWYKNVIPLLDTKKERSGKSDTFLRTAVICLVRCCVSYKDMKTCSLIYEFE
HKILNKSTLDPLLKDILDNKQELLHMDSKYGSKTTSPELAKETIEALYTTVYNHWTNA
FKLYQASLTHKPVTGKKVASVIHFIRTWRKIVKAYVSKHNNVERDLSLKNIMKONESAD
NANVLTIERMYKKLIGNSVKNTNNNSAHQMSDSEDDDDDDCEGMDVCDEASEREKK
HQESLYPINTPVTTITGDYIFKVLLELVLSPHIHPEWKIPMCDFVNRINIPKLMKAMET
DISNAVIEVRASKVNPVQILPIAANFWDFCKSGKPPSDVKFCMMFNEPSSNETLSSGA
GVFGRFIGGPFSHKSKELDIISNCLRSLLLNKEADNLSTRIWREGGSVVCFNYCPITA
RGAVLGYGEQLSERSIKALWAKKIQKQRNAADKNSRNCDLLGDEGVVSMK
TVTFGCANMLKTQNGMGKFNVVVSFEDSIQANKEGAARQYMSQQVFTHSFPALDQGK"
5185..8970
/codon_start=1
/product="ORF4"
/protein_id="AAK77673.1"
/db_xref="GI:15021396"
/translation="MASSGGFFTGIDDLFKTVIQQEKQEKNKPTQAPETEPKGPSQA
PDPVPDVPKTPTNFCPPPNPLPPPPPPKSREERLKTSKIRLNKALSDIVEATN
ERVDALKENQALNTEYDKKDNYFQVLKCSITPSVPTAIIGAHVKQVAKSSEIELAVNE
LDIKNKCSLVYNENESLKFFRDHENLILQIAVQLFSRHDNTKCVGAEICVKGNEKNKF
VNKLVVKKLPNAPSSSSTVLEIRGATRNLLENNFNKGENNTVNENKDIPSERANLDT
TKAEISHVFSTLHRLDTKRKLFFKGNTFYQRKPTFDNKFRWTEVIGWTESEAKSQITK
SLDKPTDDNLFVLPHSFRNNLADHLRLKFKNVLYKNSTAHPGKRNYYKTQETLINPQID
SAKEYKMVFAEIDKCLIDVLLAIGKNDKYTKSTVIQYRGKFRRYLIFCYAFYALNKAKH
SRAVSLPFNFFNLFSFMYCHGPFLHSASFLSTLTFVYQHMFFPMGTAAPSVASKRLM
DIDSALMKGGKGVGVRDFGSPSKTSLHRTLVSPLGFAEMAMGTMTALLSGVEVRVSP
ALQQRISKSLERWCDSVIFYFTTVLFHRFSGAKKVSLESALRLIMGQTHAHTNKVRA
AKRCRIEAAEMEGVEEEAGLTLSVAHLLGPYSIQKALGLPVPKINPLMTASSSGVN
LGDFVGVEQLLKAKREFPAEGETAGFLGWFDNLVKDSIDKYYGEGAFSDVVENNKQGM
EQNTPYDTSSALMTPIPKAFYEEKDVPQQEENSTQQRYLNRDVEEYLMASPMKMVF
VSILDKTNQKERFMSVGDIALLAVWCKRNVLKKDWNEYAIAKGNYEWLGAKMCNHLLL
ADLVNFGILGDLKITNKLDTNTDFFHRDSDRLPSVADQKKFIKNTSLSDRKQLALVHS
CVNVSTRTHVGRVTATSWAVDALRTYTRGDKDMFAALSSSLDMYHLGHTNSANFVPYF
SRNYLCNEQENGLWGYTRRTSEKLAKEELGRGRLGGLNKVGVAKTELAAAIAISSAL
DMGEVEAVMDDSSKVRKIASTCLNVNAAKVSAAREKAREASIKRLLATNAPAAGSSR
NSNRFLLKDLWGFFSDPDRKQKLIKGEAVSVLCPNTGFLHAAVPDFVIEYSFESETSI
VRLRLRLIKPEKQDEMVCPSTAPEANKGRKLVRNNQDAVLTLDDEDNIVKYKYDMVE
DEEARERLRHQDKQSVIAARISKVCERKQNPKKGRRLEDEPELQSVDEQLIRELAAIAY"
9056..10879
/codon_start=1
/product="ORF5"
/protein_id="AAK77674.1"
/db_xref="GI:15021397"
/translation="MRDDTIFNQETAVKLVRWYTEYDCCCPLVNRMERLLGSFGGGVDA
TSVRSRPALYEEDKKGDKCIPFRITSLIEGILLERALTKPDLAAAAFDVSEKLVYCSC
NNTQGNFDTVSMTIWIDGNNSKKYEVTCPSCTVEKISGGAESIHKKPMSLLAFNNLV
EKEAFAERIELKKLYLSLLTGSAAGGGMYKDSSQQSSFNGSWTSLLFHTSKKDKTRL
EAEVLVSNKIKHTSRIQPRCVCSDLLYALCSTTNSASTAYKARNLCVIEGGEFLYFK
YTIFRENGPFDSKTDLQSLVNNEPVSETNSSALAASSSLEDDDCDDDDDDDDEF
EKTKKKQPKKQTKKQKTTTSTLPPISKTNHDNMLMNVLKKGAVNGKRKWMDSLSGKKG
QHSKKLKTSAAAGGGASSDVVAGENEEENNPSSVSPTNNRDRKDYVLPCPQIEEVTIF
SQHRMNNNKLAESVVKHSVVINGNCLNLFVVNMCLNLFVYMGNFVNK
FRILTGVSCFFDRIEIVFSDQDSDSVVLSNNAAHSAILRLLSYIRENSLKRSVRTASVK
GIDFVVKSQDTNIGIPLSNKEIRERQLCSASTLSMLAGLGK"
10834..11236
/codon_start=1
/product="ORF6"
/protein_id="AAK77675.1"
/db_xref="GI:15021398"
/translation="MLSFNPEYASWFGKMITDPGVILPVSKDVVLFGSRGGSDVGIMT
LDPHDLDIKITSKRLGVEERLAQYNTLPMDFTRAMEKELNNSRNMKESIFTGIFLDTG

```
CDS            20837. .21358

Alignment Scores:
Pred. No.:                        Length:           292967
Score:                  15.5      Matches:          99
Percent Similarity:     177.00    Conservative:     67
Best Local Similarity:  35.39%    Mismatches:       165
Query Match:            21.11%    Indels:           138
DB:                     5.38%     Gaps:             20
                        14

US-10-736-868-2 (1-643) x AF369029 (1-292967)

Qy    176 ProGlyLeuValAlaAsnMetIleAlaGlyLysAsnProPheLysMetProGlnGlnMet 195
              |||.|||  |||::||||                 :::::
Db 232187 CCTGCGTTATTGGGCTAGCGGACCAACAAGAAGATGCCGACAATTATACCTGAAAACCAG 232128

Qy    196 ArgLysAla-------GlnAlaAlaProSerSerValPheGlnGlnAlaLeuAlaGln 212
              |||:::                   ::|||:||||
Db 232127 TGGAAAGTATAGTGAACAACCCCCTCGCCCAGTAGGGCCAAATAGGCAAGTTCTTCCCATG 232068

Qy    213 ArgAlaAlaMetLeuGlyLysAsnAlaProValAlaGlyGly---ArgGlyGluGlnArg 231
              ||||                |||                   ||||||||||||
Db 232067 TTGAACTTTTACTGAAAATGTGAACGCCATGGCCGATCAGCAGGTGAAGAACAGAAA 232008

Qy    232 MetMetMetAsnArgValAspGlnArgMetGlnGlnArgGluLeuGlnGluGluAspGlu 251
              ||||     :::                   :::|||||  :::||||||
Db 232007 --------AACAAGAGATGATAAC---CAACAAATAGAACCCTAGAAGAGGAAGAG 231960

Qy    252 AspAspAspAspLeuGluAspGluAlaMetAsnAspGluAlaSerSerAspGlyGluPro 271
              ||||||                        :::||||
Db 231959 GATGAAGAAGACGAGAGAACACAGGAGGAG----------GAAGAGGAAGAAGAAGAA 231912

Qy    272 GlnSerGluAlaGluHisGlnArgArgAspLeuAlaArgArgLeuLysSerSerProArg 291
              |||                   :::::::
Db 231911 CAGGAGGACGAAGAAGGAAGAGGAAGAAAAGAACCTATAGAACAAGAAAAAATGAACCCGAGAAG 231852

Qy    292 LeuLysGluLeuGluLeuGlnAsnAlaGluValGlnSerLeuLeuSerTyrGlnArgMetArg 311
              |||   |||:::         |||                   |||:::
Db 231851 GATGAAGATGCAATAGAAAATGAAAGCGTGCACTCT------------CATAGAGTTGAA 231804

Qy    312 AspSerProLeuSerLysArgArgProLeuAlaMetAsnAspGluAlaSerAlaPhe 331
              |||:::|||   :::                   :::|||
Db 231803 TCCAGCCCTATGAGTGAA----------------GGAGTAATGATGGAAATGGAATTATTTC 231756

Qy    332 ArgAlaMetGluAlaLysArgAlaLysLeuAspLysSerGlnLeuValLeuGlyLeuHis 351
              ||||||                   :::|||||
Db 231755 TTCTTCCAATAGCAGGTGCGGCAATGATAATGAAGAGGAT----------- 231714

Qy    352 GlyPheGlyGluSerAspSerAspGluAspGluAsnLeuIleAspProSer 371
              |||   :::||||::::||||                   :::|||
Db 231713 ---------GAAGAAGAGGATGAAGAAGAGGAGGAGGAGGAGGAACCTGCA 231663

Qy    372 GluAsnSerPheArgAlaProLeuArgLeuSer---SerGlyPheValGluLysLysLeu 390
              :::|||                   :::|||
Db 231662 CAAAAATCTGAAGAGCCATCTAGAAACTAGAAGAGTTCTGTTCAATCTCACACGGAATATATT 231603

Qy    391 LysSerAsnAspLeuLysSerAlaLeuAspSerArgIleLysTyrArgValAspAspVal 410
              :::                   :::|||||
Db 231602 GAAGAAGAGGAAGAGATATGAAGAATATGAAGATGAATCTCGGCACACACTCGAGGATGAA 231543

Qy    411 GlyLysTyrLeuAlaProLysProMetGluPheAsnProLysProGlyTyrTyrPhe 430
              :::||||||                   :::
Db 231542 GAA-------------ATAAGTACTATGCATCAATTTAACAATGCGCCCAGGGTTCGAAGATCT 231492

Qy    431 AlaProArgLysIle------------------------------------ 435
              :::
Db 231491 CCCCCCTCCAGATATACAAGAGTGCGAGGATGCAGTTGTGTTCCCATCATGAAGAGAG 231432

Qy    436 -------------------------ProThrArgProLysMetLeuPro 444
                                       |||   :::
Db 231431 ACAGATATTCTCCCCAGATAAAGGAACCTTCACCAAAAGCACCTAGAATGTTTTCTATT 231372

Qy    445 LeuLeuIleGlySerAsp-----------------ProLysVal 453
```

CDS

```
13311. .13982
/codon_start=1
/product="ORF7"
/protein_id="AAK77676.1"
/db_xref="GI:15021399"
/translation="MGVQKNILVGGGGGVSLLLGVTLLGVTVTEGAPAVPPSSSSYS
FTPESSVFWVEGNRVLSGTKDTLINVLGKKIPYYANSIFRHDCSETRSIQWPETSPL
GLNLIFCSCASHEHQHRTHETTEPDDLLWDGSRKTTTIILPKKWMSDTNSLWRDND
QKCGCGQAFVSSFTSTQKEVGQEWLAAHTNGKTSEGDTNSAYLPISLNELHSSLSSLT
SQKII"
```

CDS

```
13979.  .14890
/codon_start=1
/product="ORF8"
/protein_id="AAK77677.1"
/db_xref="GI:15021400"
/translation="MMMGRMSGTPMNPKDMTYFVNDFSDDIGSTPQCLVSNSDILNKR
EEWIAIWGVADSKDLLTKHQLGEREYGSEGRRRNPGVEEEEERVEEEEVEVALPYI
KKSGKLIGPRRRPLTTTTTTTTTTNPIVREVVEDFDYESFNEPEIFGSNSKLPFIR
FLDQKNWRLGIMSRVSSSIANFKIEQESSKALFCLAVWVGDEHTPKFRLSYWKNWKPF
TSAPIIVQNVGYSSDVFWHETLRSKIVDRSRDLIETKVTKKIGEDWANKKQTVVAMPI
SGIVCITVTVISIFSIVIYKKMPKP"
```

CDS

```
complement(14923. .20733)
/codon_start=1
/product="ORF9"
/protein_id="AAK77678.1"
/db_xref="GI:15021401"
/translation="MGSKRPCSSGQSPVTKKQKKNNNNNNSNPVPVINIKSYPFLATRT
QVLRSAVAAAASPSGGSSSSSSSSAVKLPDTCKEARKVLSTVSLQQSLAVRYLCNS
ISVSYAGGGISVFHLGGLPGAGKTTMVKELIAVLNDHGLIDSGSADMLLCCKSNSAKE
SLMCACKKPGGSSLMYPESVFSTLNKGFEIPVIFRKDEITLEKIQFVADKLKWKVIQV
LANLRFLVIDEYTMASCRELVFIDAVLRIAKGRPDIPFGGVFVILLGDNRQNSAVVED
NTNHIQKKIKNPSEEEKPQKNNNNKNNKKGKGEEEBEGDENEEEEGEEEBEEEE
EESDDEAETKKEEKSTFFQGSVEQDNFGQEDNAKLYTEVFIKILKMFCSRDFFGNPS
NLRNIVNKRHEAILMKSNNVKSVNNNLVSSAIKVEDCGNASNKEVTAPSSPAQSTA
EENCDEPDDEEDDDFPNNEAFLKLMERNALRASGALNGPSRLRCKSISDANEKIRS
GTTSVSDKKSSLDMMKSLPLSALIEEGICSELAHISELKKMSNANLEKYTENVCSIVF
DMMAKAMREIDYSGREKLYIVSSLSERFKDTHLTSLMDEEILNVKYVHGSDPKCIDAV
PFNSAHNRASAVAACVRNAFPRDGKDFVDETPIANYFKDNLRTVASFLENETLTYKEL
LAKSENIRSILLKKETGNNNSASSRTAAAAAYYEDDEDYCYFDEEEAMDLEDGSGCGS
GMKSSGGGGDDDDEESGEMIYRTDIPDKLHRDASTLDRVGHLVDPHVVWKKWLTENKP
SDLVRARVWYLYTLVRMQQVKPDNGKLGKSALSGRLFHSPSEWATSTGVGVGGGG
GAGADKPLHDEYWLRVLSMPISTGGDVGKSMLLPAYSSYLSALSRTYIMSSLKRIDII
KHAYSLMYGISLFDMTANLQDLVDTRMAGRSSRNGSVFMDNFDPVQYFDNIFPSMVNE
FLMYRKEDVFNNGQMMEGVKGSLKLSRVLQTAHTENNNINNRHNNSLKYSEKSIVLAM
QMVTSISKGNERRKKIEEFITKEQGQPKDMCERLMANSKAKQEKDAISSKTDKMGAI
TLTKKFVLKNAVSNLVDTSIIKETKQNNNNSSSSSTSLAAAAVENSVPALRVEVKF
VVLNMDLSDISHEKTISHKYRQQLINAIKTRSTPLFDKFTDRKILRAAESPRALTTIL
LDEKKKVTRAKSITLYQGQNVIFTTSNRMIHGTQERFVTKDTGVVTNLMYKNGELTVE
VYVERLGQKCLEIKEGRQIIGNPNIKNGGFGNNVVQYLPPESSQAMTIYSCQGHTFP
RDTIVDLSGASTQDAYAVTRNSNGQFLIIQNHSVERGNLCNIKCAMSKDKAYTMPI
GGIADFNGSDFINHDTVSVSREVAESSAAMDDDYNGDGGVTMYSAYDPSKDVVRAAEE
FILSRSGKSLSFNASWMANTAKVIQQHGLETELKNIRDFFGVNNGDVACHYEKLCNK
KMIELYTAIVRSITHYSIASGIVKQPSSKLCEEYETKQKNKDYIKIHPVFVMRAK
STIEMLLFDIAPHNKATIVFQFYVHYIFLVVEKLNVLNSSFAFLPSNPCLNQYVRPK
SITTNSTHVPNLGYESKDFAHCKDGGERDVKLRLPITSADEFSNNIEGILKKVSDTSN
QNKVNKYMDVVCKSMQHNLRRTGKFCRPTETCGLSKHGSIVTSCTCAQEKGENIHVDA
EKGWLCMSDEANVYCMLFMSKIAAASGVSEFPIKDKSILESNPETPSDTISLLAPRK
TISPTNNLHFSMSEDVLFCGQVHPMKRVQFSLHVRTGGALKSTPEEEGLPTKIFSP
NFATYPLFKKCKMYGAIIIAMTEMQGHEFAKYSTLDIRKSMFTGVGTVVDLEKISGEG
NEVMDKVDKFIVKNVSNILFKEQGKRVSFFVSCAIH"
```

```
Db  231371 CTTGGTAGTGGAGGTGAGGAACAATATGACCAACTAAACGATATAGCACCGCCTGTT
               |||                    |||         :::         |||
Qy     196 ArgLysAla-------GlnAlaAlaProSerValPheGlnGlnAlaLeuAlaGln 212

Qy     454 GlnGluGlyIleArgArgHisProSerThrGluTrpLysIleAlaLysGluSerArg--- 472
               :::                    |||       |||||||||||
Db  183331 TGGAAAGTATAGTGAACACCCCCTGCCAGTAGGGCAAATAGGCAGTTCTTCCCATG 183272

Db  231311 CCATCTATCGTGACTTTTCCTCCAGACAATGAA------ATGGGAGGAAGAAGTAGAGAC 231258
                                      ||||||||
Qy     213 ArgAlaMetLeuGlyLysAsnAlaProValAlaGlyGly-  ArgGlyGluGlnArg 231

Qy     472 -------------------------------------------------------------- 472
Db  183271 TTGAACTTTTACTTGAAAATGTGAACGCCATGGGCGGATCAGCAGTGAAGAACAGAAG 183212

Db  231257 ATCATGGACCAAGATTCGATGCTGATGCCTCCTCCTCCACCTCCACCACCACCT 231198
                                                 |||        :::
Qy     232 MetMetMetAsnArgValAspGlnArgMetGlnArgGluLeuGlnLysGluAspGlu 251

Qy     472 -------------------------------------------------------------- 472
Db  183211 ------AACAAAGAAGATGATAAC---CAACAAATAGAACCCGTAGAAGAGGAGGAG 183164

Db  231197 CATCAACCTCCTCAACTCAAACTACCAATATTCTTCCTCCCCCTCCACCTCCACCT 231138
                                      ||| :::      |||||||||||||
Qy     252 AspAspAspLeuGlyLysGluGluGlyValProArgArgSerSerAspGlyGluPro 271

Qy     473 -------------ValLeuThrAsnLeuLysAsnAsnProSerLeuAlaAlaLeuPheMet 488
                    :::|||   |||
Db  183163 ------GAAGAGGAAGAGGAACAGACAGGAGAG 183116

Db  231137 ACAAATCAGTCCTTATTTCAAACAACAACAATCCATCCTTTTGAGTACAGTGGTT 231078
                                      :::  |||||||||||
Qy     272 GlnSerGluAlaGluHisGlnArgArgAspLeuAlaArgLeuLysSerSerProArg 291

Qy     489 AspAspLysLeuGluAsnThrLeuLysGlyArgGlnMetLeuThrArgAspGluGlnLysGlyGly 508
                :::|||||||||||||
Db  183115 CAGGAGGAGGAGAAGAAAAGAACCTATAGAACAAGAAAAAAATGAACCCGAGAAG 183056

Db  231077 --GGAAAGGTTAACACACATTAGGTGGGGAAGAGAGGCCGAAGAAAGATTACACAAGACT 231021
                                      |||       :::
Qy     292 LeuLysGluGluLeuGlnLeuAsnAlaGlyValGlnGlnLeuSerLeuLeuSerTyrGlnArgMetArg 311

Qy     509 -------ArgThrArgValLysThrArgIleAlaArgAlaLeuProArgLeuPhe 522
                      :::|||:::|||
Db  183055 GATGAAGGATGCAATAGAAAATGAAAGCGTGCACTCT---CATAGAGTTGAA 183008

Db  231020 ATGGAGTCTATAATCTTAAAAACAAGGGTAAAAACTCTT------------------- 230982
                                      |||:::|||
Qy     312 AspSerProLeuSerLysArgArgProLeuAlaMetAsnAspGluSerAlaPhe 331

Qy     523 GlyAlaProThrAlaLys-------------AlaGluMetIleAspAlaLysValPheGln 538
                   |||       :::|||||||
Db  183007 TCCAGCCCTATGAGTGAA--------GGAGGTAATGATGATGAATGATTATTTC 182960

Db  230981 ---CTAGAACAACAAAGAACCTCTAATGCCTCTGACTCGTGGTGAAAGTT---GTGTTCCAA 230928
                                      |||::::::
Qy     332 ArgAlaMetGluAlaArgAlaLysLysAspAlaLeuAspGlnLeuLeuValLeuGlyLeuHis 351

Qy     539 AspIleGluGluArgProIleProPro 547
                |||   |||||| |||
Db  230927 GAT---CCAGAAACCCAGTAAAACCA 230904

Qy     352 GlyPheGlyGlyValSerAspAspAspLysGluGluAspAlaGluGluAspGlyGluGluAsnLeuIleAspProSer 371
                                      :::|||||||:::
Db  182917 --------GAAGAGGATGAAGAGGAAGAGGAGGAGGAACCTGCA 182867

Qy     372 GluAsnSerPheArgArgAlaProLeuArgLeuSer---SerGlyPheValGluLysLeu 390
                      :::|||   |||
Db  182866 CAAAAATCTGAAGAGATAGAAACTAAAGATGTCTGTTCAATCTCACACGGAATATATT 182807

Qy     391 LysSerAsnAspGluLeuLysSerAlaLeuAspAspValIleTyrTyrArgValAspAspVal 410
                |||       :::  |||       |||:::|||
Db  182806 GAAGAAGAGGAAGAATATGAAGAATATGAAGATGAATCTCGGCACACACTCGAGGATGAA 182747

Qy     411 GluLysTyrLeuAlaProLysProMetGlyPheAsnProLysProGlyTyrTyrPhe 430
                |||:::|||
Db  182746 GAA------ATAAGTACTATGCATCATCAATTTAACAATGCCCCAGGGTTCGAAGATCT 182696

Qy     431 AlaProArgLysIle 435
                |||
Db  182695 CCCCCTCCAGATATACAAGAGTGCGAGGATGCGAGTTGTTCCTCCAATCATGAAAGAG 182636

Qy     436 -------------------------ProThrArgProLysMetLeuPro 444
Db  182635 ACAGATATTCTCCCCAGATAAAGAACCTTCACCAAAAGACCTAGAATGTTTTCATT 182576

Qy     445 LeuLeuIleGlySerAsp----------------ProLysVal 453
                |||   :::
Db  182575 CTTGGTAGTGGAGGTGAGGAACAATATGACCAACTAAACGATATAGCACCACCGCCTGTT 182516

Qy     454 GlnGluGlyIleArgArgHisProSerThrGluTrpLysIleAlaLysGluSerArg--- 472
                :::                    |||       |||||||||||
Db  182515 CCATCTATCGTGACTTTTCCTCCAGACAATGAA------ATGGGAGGAGAAGTAGAGAC 182462

Qy     472 -------------------------------------------------------------- 472
Db  182461 ATCATGGACCAAGATTCGATGCTGATGCCTCCTCCTCCACCTCCACCACCACCT 182402

Qy     472 -------------------------------------------------------------- 472
Db  182401 CATCAACCTCCTCAACTCAAACTACCAATATTCTTCCTCCCCCTCCACCTCCACCT 182342
```

RESULT 12
AX151396/c

```
LOCUS       AX151396    305107 bp    DNA    linear    PAT 22-JUN-2001
DEFINITION  Sequence 1 from Patent WO0138351.
ACCESSION   AX151396
VERSION     AX151396.1    GI:14533479
KEYWORDS    .
SOURCE      Shrimp white spot syndrome virus
  ORGANISM  Shrimp white spot syndrome virus
            Viruses; dsDNA viruses, no RNA stage; Nimaviridae; Whispovirus.
REFERENCE   1
  AUTHORS   Xu,X., Yang,F., He,J., Pham,L.Z., He,M., Ye,Y., Shen,Y. and
            Kodira,C.
  TITLE     Nucleotide sequence of the shrimp white spot syndrome bacilliform
            virus (wsbv), systems containing this sequence and detection kits
  JOURNAL   Patent: WO 0138351-A 1 31-MAY-2001;
            PE Corporation (NY) (US) ; The Third Institute of Oceanography,
            State Oceanic Administration (CN) ; Sinogenomax Co., Ltd. (CN)
FEATURES             Location/Qualifiers
     source          1..305107
                     /organism="Shrimp white spot syndrome virus"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:92652"
ORIGIN
```

Alignment Scores:
```
Pred. No.:                16.4              Length:        305107
Score:                    177.00            Matches:       99
Percent Similarity:       35.39%            Conservative:  67
Best Local Similarity:    21.11%            Mismatches:    165
Query Match:              5.38%             Indels:        138
DB:                       6                 Gaps:          20
```

US-10-736-868-2 (1-643) x AX151396 (1-305107)

```
Qy     176 ProGlyLeuValAlaAsnMetIleAlaGlyTyrLysAsnProPheLysMetProGlnMet 195
               :::                    |||::::::         |||
Db  183391 CCTGCGGTTATTGGCGGTTAGCGCGACCACAACAAGAAGATGCGACAATTATACCTGAAAACCAG 183332
```

```
Qy    473 --------------ValLeuThrAsnLeuLysAsnAsnProSerLeuAlaAlaLeuPheMet 488
                        :::  ||||||||||||||||      :::         :::
Db 182341 ACAAATCAGTCCTTATTTTCAAACAACAACAACAATCCATCCTTTTCAGTCACAGTGGTT 182282

Qy    489 AspAspLysLeuGluAsnThrArgLeuLysGlyArgGlnMetLeuThrAspGluGlnLysGly 508
                ||||      |||||||||||||| |||:::::
Db 182281 ---GGAAAGGTAACAACACCATTAGGTGGAAAGAGGCCGAAGAAAGATTACCAACAAGACT 182225

Qy    509 ---------------ArgThrArgValLysThrIleArgAlaLeuProArgLeuPhe 522
                              :::||||||||||||
Db 182224 ATGGAGTCTATAATCTTAAAAACAAGGGTAAAAACTCTT------------------ 182186

Qy    523 GlyAlaProThrAlaLys-----------AlaGluMetIleAspAlaLysValPheGln 538
                         |||              :::||||:::::   |||||||
Db 182185 --CTAGAAACAACAAAGAACCTTCAATGCTCTGAACTGGTGAAAGTT---GTGTTCCAA 182132

Qy    539 AspIleGluArgProIleProPro 547
                |||           |||
Db 182131 GAT---CCAGAAAACCCAGTAAAACCA 182108

RESULT 13
AF332093/c
LOCUS       AF332093             305107 bp    DNA     circular VRL 20-NOV-2001
DEFINITION  White spot syndrome virus, complete genome.
ACCESSION   AF332093
VERSION     AF332093.1  GI:17016399
KEYWORDS    .
SOURCE      Shrimp white spot syndrome virus (white spot bacilliform virus)
  ORGANISM  Shrimp white spot syndrome virus
            Viruses; dsDNA viruses, no RNA stage; Nimaviridae; Whispovirus.
REFERENCE   1  (bases 1 to 305107)
  AUTHORS   Yang,F., He,J., Lin,X., Li,Q., Pan,D., Zhang,X. and Xu,X.
  TITLE     Complete genome sequence of the shrimp white spot bacilliform virus
  JOURNAL   J. Virol. 75 (23), 11811-11820 (2001)
  MEDLINE   21548311
  PUBMED    11689662
REFERENCE   2  (bases 1 to 305107)
  AUTHORS   Yang,F., He,J., Lin,X., Li,Q., Pan,D., Zhang,X. and Xu,X.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-DEC-2000) Key Laboratory of Marine Biotechnology, The
            Third Institute Oceanography, SOA, 178 Daxue Road, Xiamen, Fujian
            361005, P.R. China
FEATURES             Location/Qualifiers
     source          1..305107
                     /organism="Shrimp white spot syndrome virus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:92652"
                     /note="WSBV
                     synonym: white spot bacilliform virus"
     CDS             join(300501..305107/1..448)
                     /codon_start=1
                     /product="wsv001"
                     /protein_id="AAL33534.1"
                     /db_xref="GI:17016928"
                     /translation="MAYIDQGAFGAKSVMQQQYLLPPSNRLSKQQPPLASSSLQPSSS
                     NKPRSTSRVTDIFVVTCVLIAAFISNSFSVTKNVVKLSKEQTEKILEKDLPDKVYK
                     LFENLKDGTFGIGEDEEEREREEGEELETKNIRLKRKWKWTEOGDQGIKVRKLH
                     GPRGRERGETGPAGAVGPAGPQGERGAIGPAGKDGAVGPAGPQGERGAIGPAGKDGAVG
                     PQGPPGERGENGRPGRDGAVGPQGERGAIGPAGKDGAVGPQGERGAIGPAGKDGAVGP
                     AGPQGERGENGRPGRDGAVGPAGPGGERGAIGPAGRDGAIGPAGPQGERGAIGPAGKDGAVGP
                     GVDGSVGPQGERGEIGRPGRDGAVGPAGPQGRRGATGRAGKDGAVGPAGPQGEKGEAG
                     RGETGPVGPGRGEPGLAGLPGRDGAIGPAGPPGERGATGLPGRNGVDGSIGPQGRRGAT
                     GRAGKDGAVGPAGPGERGATGIPGRDGVDGSVGPPGERGETGPAGRDGSVGPAGPHG
                     ERGENGRPGRDGATGIPGAGPQGEKGENGRPGRDGATGIPGPRGETGAMGKNGVDGS
                     MGPQGRRGATGRAGKDGAVGPAGPPGERGAIGPAGTPGAGRDGSVGPAGPQGEGTGLTGSPGRD
                     GATGPIGPAGPQGEKGENGRPGRDGATGIPGAGPQGEKGENGRPGRDGATGPIGPAG
                     PQGETGLTGRPGRDGATGPIGPRGETGAMGKNGVDGSTGPQGRRGATGRAGKDGAVGP
                     AGPPGERGENGRPGRDGATGPIGPAGPQGETGLAGLPGRDGAIGPAGKDGAIGPAGKDGAVGP
                     GATGMGPEGERGETGPIGPAGPQGATGLPGRDGVDGSVGPQGKRGLIGRTGRDGAIG
                     PVGPAGPKGETGLAGLPGIDGKDGSVGPQGAIGPIGPRGQRGDGIMGPRGLRGEGKGAGND
                     MGPQGLRGATGAPGPGERGLAGPQGERGATGIPGRDGEMGPAGKDGIEGPRGGQDGTTG
                     GLEGPEGRDGAPGPAGPIGPQGIRGLKGIQGRPGRDGEPGPAGKDGIEGPRGQGTTG
```

```
          AKGPRGLRGFQGRTGETGGAQGSRGEKGDRGLTGPGQRGRDGPPGEGEGPQGLRGERGAPGP
          RGPRGIRGRSGPQGSNGVQGPRGPRGTKGRTGQLTGIEGPRGPRGIQGKEGRMGKI
          GHRGEKGDKGDRGEQGIAGADEKGPRGLRGRIRGPIGAPGKPGTEGVRGPRGVRGVPG
          YPGAQGELGPQGPTGPGPQGPAGPQGPMGRTGDTGPMGPPGAVGPRGEKGGRGRKGKNGP
          KGADGKDAVNIIQKYSITHARAEIMWEGNEIGEAYIGRSYGTDTIPVMIENRIGMTNE
          DKKNEYCIQVMTMHSITTRGRTSGVFVVSNKTDYILLVTLLMPESVSCRTDVSTNARS
          ERVNAVRERESKSYRFIRPSDQSIGTHSRSKIAVVMYPDASMSYSVDTLDADVARRET
          TSVLLLAETIHGEKDRGFYADRGTVGRLMVPPTEEELLVLQADPITPTIADPITVAE
          VGNPVVRIKTPSCDELNKFMRFSTNFSFLKYVKRFMAQRASSNRNDMSAFLLDLIFE
          KLIEFLLSTGWNREALGGTTGAILMIRTATCRLAQSYYPLPETVGMDFLQTPGPDYNK
          SFPNNERYIYEFQATM"
     CDS             complement(492..1118)
                     /codon_start=1
                     /note="vp24"
                     /product="wsv002"
                     /protein_id="AAL33006.1"
                     /db_xref="GI:17016400"
                     /translation="MHMWGVYAAILAGLTLILVVISIVVTNIELNKLDKKDKDAYPV
                     ESEIINLTINGVARGNHFNFVNGTLQTRNYGKVYVAGQTSDSELVKKKGDIILTSLL
                     GDGDHTLNVNKAESKELELYARVYNNTKRDITVDSVSLSPGLNATGREFSANKFVLYF
                     KPTVLKKNRINTLVFGATFDEDIDDTNRHYLLSMRFSPGNDLFKVGEK"
     CDS             571..900
                     /codon_start=1
                     /product="wsv003"
                     /protein_id="AAL33007.1"
                     /db_xref="GI:17016401"
                     /translation="MSSSNVAPNTSVLILFFFKTVGLKYNTNLLAENSLPVAPRPGDS
                     ETESTVISLFVLLYTLAYNSNSLDSALFTFSVWSPSPSKDVRIMSPFFFTSSSSDVPW
                     PATYTFP"
     CDS             complement(1197..1511)
                     /codon_start=1
                     /product="wsv004"
                     /protein_id="AAL33008.1"
                     /db_xref="GI:17016402"
                     /translation="MYLSHIRQTPLVEERRALTFKMYHHNNNQHSFVNCQCRRTSSS
                     INCSSCSRETFNSVKAIQYFNKTSRNNTAHHFKMPASKDRNYSSFEYAETAVAAHNIS
                     QW"
     CDS             1437..1664
                     /codon_start=1
                     /product="wsv005"
                     /protein_id="AAL33009.1"
                     /db_xref="GI:17016403"
                     /translation="MMVHFEGESTSFFNQRCLSNVAQVHPRSPFIPTSYSSFRSSVMA
                     LILLVIPRSTKNLISPSDLFKTFHDSIGSRV"
     CDS             complement(1538..2425)
                     /codon_start=1
                     /product="wsv006"
                     /protein_id="AAL33010.1"
                     /db_xref="GI:17016404"
                     /translation="MKNSRQRSGVWRGNSCLYKSFYFSGAIIECKKIRIIMMFLLLSL
                     ILFVCFVGVVVGVIFMSRPNKTTTTSNKKTKDKEKEKEDDTEGAVLGRRREPENRPIG
                     RDEEGAVEDGKEEEVFEFEQPSVNTGSNTGGGTGTVPGEGLLPPPPTPTPTPPT
                     PTPTPPPPTRTPSPSSSLGEDDDDIDIDFFDDNDIDEFLDSGEEMEDEEBEDLDTL
                     LSRLETGMSGEEVDFDASSAYIQPDPVVVKNIERSDYTLDPMESWKVLNRSEGDIRFF
                     VDRGITNKIKAMTEDLKEL"
     CDS             1711..1938
                     /codon_start=1
                     /product="wsv007"
                     /protein_id="AAL33011.1"
                     /db_xref="GI:17016405"
                     /translation="MQMMHQNLLLRRSCLFLVLKGVCPNLPLLRLLPFLLHYLKIRLY
                     HYHQSLCLYHHHHLPQEMKKEMGFVSGEVEE"
     CDS             complement(1749..2363)
                     /codon_start=1
                     /product="wsv008"
                     /protein_id="AAL33012.1"
                     /db_xref="GI:17016406"
                     /translation="MPVSSLERSVGKSSSSSSSSISPLSNVSISLSSKSMSISSSS
                     SSPRDEEGDGVRVGGGGGVGVGGGVGVGGGGNGKPSPGTVPPPPVLDPVFTL
                     GCSNSKTISSSFSPSSTAPSSSLPIGLFSGSLRPNTAPSVSSSFSFSLSFFVLLDVVV
                     VLFGRDMKITPTTTPTKHTKSMRDNKRNIIILIFLHSIMAPEK"
     CDS             complement(2385..2672)
                     /codon_start=1
```

```
CDS
     /product="wsv009"
     /protein_id="AAL33013.1"
     /db_xref="GI:17016407"
     /translation="MDQLTENPSLLAERPVFRQKVIDLYREEILRDLLIKKMTADMTSA
     IEEACTAAVADGSFWRDMRYEMVAHQHDVTFAAKQNMEGYEEFAAEVRRVEG"
     complement(2703..2996)
     /codon_start=1
     /product="wsv010"
     /protein_id="AAL33014.1"
     /db_xref="GI:17016408"
     /translation="MDILEDIYKSAITLVLQSPEFVNDVKQEASQVVEGLIPSIREAV
     FRRLLEEBRKKHEDEVGDVEDKRQAVIDKANTMITTMAAEYLESVDILEEFGF"
     3051..6956
     /codon_start=1
     /product="wsv011"
     /protein_id="AAL33015.1"
     /db_xref="GI:17016409"
     /translation="MWLMTSCPIYVLVGCNTSLYKAIEGVKTGHTMVLTLSCTTRRVA
     SSKGNFSKEDAVLGNQFPLLKKSNNLSIARPPSIESFSAVEKIFREWNESGGEKIFD
     ISQNEEEWMDIISLVESVYEPVFSKSLKPDKLADKTCLTAAAFAALASAVDEKLTILS
     GSDGSVLQRTTKVMKKDPKKIAESLLNNEKWTSLLLDRLKTAKKLLSRGALKSAERV
     EVLHRLNKLKEAPLPHHPSLFDNFSGGKTSAVSAGTVIASDMHFKLVEHIFKVSFRKW
     GPCGDKTESGEEEDEEEEBEEKKHSISRFVLQFMNGHNGQHYHRPESASVYFCDYYDY
     LAYRNLPNEYKLSSMHPGTFNMEDLPFRPFAVPSTYKTELEYKRFVQSTNLPQLSFDY
     GEFLCYCIFGADWYKHLGDVVDSLENSSMISFDSGTLSGVYKNTANYKRLGKKRNGIA
     DLAVRSMAEFIRTEAHKALTAEEMEEEEEEEAEEEAMQEPAEVDFLSVPHLRRKIR
     QAVSVLNNFVENDLSILVSNFKNVLTDDTVSGTDTDNFGSSGEFEALSSHLFLSRILD
     EVHILRNTDIQRTLFSTHVSLSDSPPSRVRGSNVFNNNAGNISSLQTYGGIEELPE
     NVLVGLSGGFEDTDMYSGEDVVVWDGCDGGKVLSVTFNCGDNFIQLHEKTAETFKDD
     TDLVERIRDVLQTASKTGNLNKKAYSRKNIYAVLRENGIERPGDDFTEKGIALKDKTN
     QPPPPARSAKITVEGVKGPFSGFRDILETRALITYSAETFRDLGQGIVKETEGLTAAT
     VAETSFSEGLAESLRSDANLGLEFSEDAKTVVFKNDTSRSLLEETRALRANNTSFSSF
     ARDMGVQVSADLDAEFAAEMRETYPDAALEQNLKLDLDKFEETIPESQVKKLKIDSYL
     TENPERAGKEINDTELSKATDSVLGKKLGNAVTVLMNNFGKVTIVGASVVAGPLGPA
     AVALVHASRGAHLNVVDHTSPKGVISYKIVDFSCADRNTGWAKPTKHPPREEIDHVIA
     LDASFLTENGAYVFPEDGGPKSKYKAYAPICGTKDAAQECGSWATFDDPHSVLPWVA
     SMKDLPKGQSLSCDKGMSTLKAVSSVLLSIGKDVAEAIFEVAEDAVVGLASKAISAVI
     NNPLFIFGVPLGFGIAATRLNPSNWKTGLIVFSILLVVLIVRFFAGSGPLTLNWFGA
     KNSAKRKTEQFEDGGGNRSKIVLAEKDNANSKLQSRRNETGPMRLEELPGHEDLRPV
     FFPATTNYSKSAKILGYKSKPFNDFYTKLINTDIIKMDR"
CDS  complement(3107..3358)
     /codon_start=1
     /product="wsv012"
     /protein_id="AAL33016.1"
     /db_xref="GI:17016410"
     /translation="MSKIFSPPLSFHSLNIFSTDAEKDSIEGGLAIDKLFDFFKMGNW
     FPNTASSLEKFPLLDATRLJVVQERVSTMVCFVFTPSIAL"
CDS  complement(3713..3955)
     /codon_start=1
     /product="wsv013"
     /protein_id="AAL33017.1"
     /db_xref="GI:17016411"
     /translation="MECFFSSSSSSSSSSSSSSPLSVLSPQGPHFLKETLKICSTNLKCIS
     DAMTVPADTADVFPPLKLSNRLGWWGRGASLSLFNR"
CDS  complement(4111..4335)
     /codon_start=1
     /product="wsv014"
     /protein_id="AAL33018.1"
     /db_xref="GI:17016412"
     /translation="MSENQMISWNYFLENLPHPPGVCTNLLRRYSNIKTPHSRNLAGE
     DLLIAQTFCTLILSYKLKVLRKGGKVNPPY"
CDS  complement(4574..4966)
     /codon_start=1
     /product="wsv015"
     /protein_id="AAL33019.1"
     /db_xref="GI:17016413"
     /translation="MSVSSNPPDKPTSTFSGNSSIPPYVCREEMFPALLLKITLLPRT
     RLGGDLSRDTCVENRVLCISVFLRMCTSSNILERNKWEDNASNSPLEPKLSVSVPDT
     VSSVNTFLKLETNIERSFSTKLFNTETA"
CDS  4654..4908
     /codon_start=1
     /product="wsv016"
     /protein_id="AAL33020.1"
```

```
                                        /db_xref="GI:17016414"
                                        /translation="MILYQEQIRRTLVLVENLKHYLPIYSFQEYWMKCTFLGILIYKE
                                        PYFQRTCLCRINLPLAVSVEAMSTLIITILGTFLPCKRMAV"

Alignment Scores:
Pred. No.:              16.4            Length:         305107
Score:                  177.00          Matches:        99
Percent Similarity:     35.39%          Conservative:   67
Best Local Similarity:  21.11%          Mismatches:     165
Query Match:            5.38%           Indels:         138
DB:                     14              Gaps:           20

US-10-736-868-2 (1-643) x AF332093 (1-305107)

Qy   176 ProGlyLeuValAlaAsnMetIleAlaGlyLysAsnProPheLysMetProGlnGlnMet   195
             |||       |||:|||                        |||:::::|||
Db 183391 CCTGCGTTATTGCGCTAGCGCACAACAAGAGATGCGACAATTATACCTGAAAACCAG 183332

Qy   196 ArgLysAla-------GlnAlaAlaProSerSerValPheGlnGlnAlaLeuAlaGln   212
             |||:::           :::|||
Db 183331 TGGAAAGTATAGTGAACACCCCTGCGCCAGTAGGCGCCAAATAGGCAAGTTCTTCCCATG 183272

Qy   213 ArgAlaMetLeuGlyLysAsnAlaProValAlaGlyGly--ArgGlyGluGlnArg     231
             |||   :::|||         :::                |||:::::||:::
Db 183271 TTGAACTTTTTTACTTGAAAATTGAACGCAGCCATGGCCGGATCAGCAGTGAACAACAGAAG 183212

Qy   232 MetMetMetAsnArgValAspGlnArgMetGlnIleArgLeuGlnGluGlnAspGlu     251
                                 --------AACAAAGAAGATGATAAC--
Db 183211 ------------AACAAAGAAGATGATAAC--CAACAAATAGAACCCTAGAAGGAAGAG 183164

Qy   252 AspAspAspGluLeuGlnLeuAspValProArgArgArgSerSerAspGlyGluPro     271
             |||:::|||   :::                     :::
Db 183163 GATGAAGAAGAGGAAGAGAACAGGAGGAG-----GAAGAGAAGAGGAAGAAGAAGAA 183116

Qy   272 GlnSerGlyAlaGluHisGlnIleArgArgAspLeuAlaArgArgLeuLysSerProArg   291
             |||   |||:::           :::                :::
Db 183115 CAGGAGGAGGAGGAAGAGAAGAAGAAACACCTATAGAACAAGAAAAAAATGAACCCGAGAAG 183056

Qy   292 LeuLysGluLeuGluLeuGlnAsnAlaAlaGluValGlnSerLeuLeuSerTyrGlnArgMetArg 311
                                                            --------CATAGAGTTGAA
Db 183055 GATGAAGAGTGCAATAGAAAATGAAAGCGTGCACTCT-------CATAGAGTTGAA 183008

Qy   312 AspSerProLeuSerLysArgGluProLeuAlaMetAsnAspGluAlaGluSerAlaPhe   331
             |||:::|||:::||:::|||:::
Db 183007 TCCAGCCCTATGAGTGAA-------GGAGGTAATGATGATGGAATGGATTATTTC 182960

Qy   332 ArgAlaMetGluAlaArgAlaLysLeuAspArgIleLysSerGlnLeuValLeuGlyLeuHis  351
             :::   :::
Db 182959 TTCTCTTCAATAGCAGTGGCGCAATGATAATGAAGAGGAT----------- 182918

Qy   352 GlyPheGlyGluSerAspAspAspAspGluGluAspGluAsnLeuIleAspProSer     371
             :::|||         ||:::||:::|||:::
Db 182917 --------GAAGAAGAGGATGAAGAGGAGGAGGAAGAGGAGGAGGAACCTGCA 182867

Qy   372 GluAsnSerPheArgAlaProLeuArgLeuSer---SerGlyPheValGluLysLeu     390
             :::                                :::
Db 182866 CAAAAATCTGAGAGACGCATAGAAACTAAAGAGTCTGTTCAATCTCACACGGAATATATT 182807

Qy   391 LysSerAsnAspGluLeuLysSerAlaLeuAspArgIleLysTyrArgValAspAspVal   410
             :::   :::|||:::         :::
Db 182806 GAAGAAGAGGAAGAATATGAAGAGATGAAGAATCTCGGCACACACTCGAGGATGAA 182747

Qy   411 GluLysTyrLeuAlaProLysProMetGluPheAsnProLysProGlnProGlyTyrPhe   430
             |||         ||:::||:::|||:::
Db 182746 GAA--------ATAAGTACTATGCATCAATTTAACAATGCGCCCAGGGTTCGAAGATCT 182696

Qy   431 AlaProArgLysIle                                             435
             |||
Db 182695 CCCCCTCCAGATATACAAGAGTGCGAGGATGCGCAGTTGTTCCTCCAATCATGAAAGAG 182636

Qy   436 ProThrArgProArgLysMetLeuPro                                 444
                                 :::
Db 182635 ACAGATATTCTCCCCAGATAAAGGAACCTTCACCAAAAGCACTAGAATGTTTCTATT 182576
```

Qy   12 GlyPheCysIleAlaProLeuSer-------------------ProLysVal 27
                ||| ||| |||
Db   3999 GGCTTCTCAGCCAGTGAGGTGAGGAACAATATGACCAACTAAACGATATAGCACCACCGCCCTGG 4058

Qy   28 AlaProGly------AlaLeuLeuSerSerLeuValGlyLysHisGlnLysLeuPro 45
             :::|||||| ::|||||||||| |:: :||||||
Db   4059 AGTCCCAGTGCAGACACTCAGGAGCTG-CTGCAGGAGGAGGAACCGGCAGAAGCTGAGC 4117

Qy   46 LeuAlaProSerMetGluAlaLeuGlu------------------- 54
             |||:::   :::||| :::|||
Db   4118 CTGGCACCAAGCTTCAAGCAGGTGGAGGACGAGGAGAATTCCTTCCGGGAGCAGCTGGAG 4177

Qy   55 --------------------------LeuMetGlyGlyValGlnPheVal 61
                                       ::: |||
Db   4178 GAGGAGGAGGCCAAGCACAACCTGGAGAAGCAGATCGCCACCCTCCATGCCCAGTGGCC 4237

Qy   62 AspAlaLeuIleLysGlyGlnMetGluMetAlaLysGlyAlaPheLysThrGlnLeuLeu 81
             ||| |||||||||:: ::|||||||::: :: :::|||
Db   4238 GAC------ATGAAAAGACGATGGAGGACAGTGTGGGGTGCCTGGAAACTGCTGAG 4288

Qy   82 GluValLeuGluLysValHisProAsp------------------- 90
             ||||| |||||||
Db   4289 GAGGTGAAGAGAGGACACCTGGAGGGCTGAGCCAGCGGCACGGAGGAGAAG 4348

Qy   91 --------GlnPheAspLysTyrLysLeuLys-------------------ValAspAsp 102
                      ||||| ::|||:::||| :: |||
Db   4349 GTGGCCGCCTACGACAAGCTGGAGAAGCAAGACGCGGCTGCAGGAGCTGCACGGAGCGAC 4408

Qy   103 LeuAlaLaAAspAlaValMetGlnGlnAlaGluLeuGlnProLysSerGly 122
             ||| |:::::||||||
Db   4409 CTGCTGGTGACCTGGACCACCAGCGCCAGAGGCGTGCAACCTGGAGAAGAACAGCAGAAG 4468

Qy   123 AsnAlaPheIleAspMetLeuAsnGlyAsnGlyIleProIleGly-------- 137
             ||| |||
Db   4469 AAG------TTTGACCAGCTCCTGGCGGCGGAGGCCCAGAGAGGACCATCTCTGCCAAGTATGCAGAG 4522

Qy   138 --------------------------SerSerIle 140
                                       |||
Db   4523 GAGCGCGACCGGGCTGAGCGCGGAGGCCCAGAGGAGGAGGACCAAGCGCTCTGTCGCTGGCC 4582

Qy   141 ArgGlyLeuGluAspAlaIleArgThrAspProSerGlu 160
             ||||||::|||:::|||| :::::::::::||
Db   4583 CGGGCCCTGGAGGAAGCCATGAGCAGGGACTGAGCTGAGCGGCTCAAC- 4633

Qy   161 GlnIleAlaLysAlaValMetAspLysPheGlnThrGlnIleLeuProGlyLysLeuValAlaAla 180
               :::||:|||||||| :::
Db   4634 AAGCAGTTCCGCACGGAG-----------------ATGGAG 4657

Qy   181 AsnMetIleAlaGlyLysAsnProPhe---LysMetProGlnGlnMetArgLysAlaGln 199
             ::::::::::::: ||| |||
Db   4658 GACCTTATGAGCTCCAAGATGATGTGGGCAGCTGTGTCCAGAGCTGGAGAAGTC-CAA 4716

Qy   200 AlaAlaProSerSerValPheGlnGlnAlaLeuAlaLeuAlaGlnAlaArgAlaMetLeuGlyLysAsn 219
             ||||| |||||||
Db   4717 GCGGGGCCT----------------AGAGCAGCAGGTGAGGAGGAGAT 4746

Qy   220 -------------AlaProValAlaGlyAlaGlyArgGlyGluGlnLysArgMetMetAsn-Ar 236
                          ||||||| ||||||||
Db   4747 GAAGAGCAGCAGCTGGAAGACTGGAGGAGACGAGCTGCACCGAAGATGCCAAGCTGCG 4806

Qy   236 gValAspGlnArgMetGln------------QlnArgGluLeuGlnGluGluAs 250
             ::||| |||
Db   4807 GTTGGAGGTCAACCTGCCCATGAAGGCCCAGTTCGAGCTGCGCGGGGCCGGGGA 4866

Qy   250 pGluAspAspAspAspLeuGluAspGluAsnGluGlnValProArgArgSerSerAspGlyGl 270
             |||| |||||||
Db   4867 CCAGCAGCAGCGACGGAGAAGAAGCAGCTGGTTCAGCAGGTGCGGGATGGAGGCGCAGA 4926

Qy   270 uProGlnSerGluGlnAlaGluHisGlnAlaArgGluAlaArgLeuLysSerSerSerPr 290
             ||| :::   :::   ::::::|||
Db   4927 GCTGGAGGACGAGGAGGAGGAGCAGCAGCTGGTCGGCGCCGGCGGCGCGAGAAGAAGCTGGGAT 4986

Qy   445 LeuLeuIleGlySerAsp-------------------ProLysVal 453
             ||| ||| |||
Db 182575 CTTGGTAGTGGAGGTGAGGAACAATATGACCAACTAAACGATATAGCACCACCGCCCTGTT 182516

Qy   454 GlnGluGluIleArgArgHisProSerThrArgLysIleTrpLysIleAlaLysGluSerArg-- 472
             ::: ||| ||| :: ::|||||||||
Db 182515 CCATCTATCGTGACTTTTCCTCCAGACAATGAA------ATGGGAGGAGAAAGTAGAGAGAC 182462

Qy   472 ------------------- 472
Db 182461 ATCATGGACCAAGATTCGATGCTGATGCCTCCTCCTCCTCCACCTCCACCACCACCT 182402

Qy   472 ------------------- 472
Db 182401 CATCAACCTCCTCAACTCAAACTTACCCATATTCTTCCTCCCCCTCCACCTCCACCT 182342

Qy   473 ------ValLeuThrAsnLeuLysAsnAsnProSerLeuAlaAlaLeuPheMet 488
                   ::: :::||||||||||||||| ::: :::|||
Db 182341 ACAAATCAGTCCTTATTTCAAACAACAACAACAATCCATCCTTTTGAGTACAGTGTT 182282

Qy   489 AspAspLysLeuGluAsnThrLeuLysGlyArgGlyMetLeuThrAspGluGlnLysGly 508
             :::|| ||||||:::|||
Db 182281 --GGAAAGGTTAACACACATTAGGTGGAAGAGGCCGAAGAAGATTACACAAGACT 182225

Qy   509 -----ArgThrArgValLysThrIleArgAlaArgLeuProArgLeuPhe 522
                  |||||||||::: ||||||:::|||
Db 182224 ATGGAGTCTATAATCTTAAAACAAGGTAAAAACTCTT----------- 182186

Qy   523 GlyAlaProThrAlaLys----------AlaGluMetIleAspAlaLysValPheGln 538
             ||| ::|||||||
Db 182185 ---CTAGAAACACAAAGAACCTTCAATGCTCTGAACTGGTGAAAGTT---GTGTTCCAA 182132

Qy   539 AspIleGluGluArgProIleProPro 547
             ||| ::||| |||
Db 182131 GAT---CCAGAAAACCCAGTAAAACCA 182108

RESULT 14
AX686226
LOCUS   AX686226   7396 bp   DNA   linear   PAT 29-MAR-2003
DEFINITION   Sequence 35 from Patent WO02059315.
ACCESSION   AX686226
VERSION   AX686226.1 GI:29372043
KEYWORDS   .
SOURCE   Homo sapiens (human)
ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
AUTHORS   Shinkets,R.A., Patturajan,M., Vernet,C.A., Casman,S.J.,
             Malvankar,U., Shenoy,S., Spytek,K.A., Gangolli,E., Miller,C.,
             Boldog,F., Li,L., Taupier,R.J., Kekuda,R., Smithson,G.,
             Zerhusen,B.D., Liu,X., Colman,S.D., Tcherney,V., Si,J., Edinger,S.,
             Stone,D., Sciore,P., Millet,I. and Rothenberg,M.
TITLE   Human nucleic acids and polypeptides and methods of use thereof
JOURNAL   Patent: WO 02059315-A 35 01-AUG-2002;
             Curagen Corporation (US)
FEATURES   Location/Qualifiers
source   1..7396
                     /organism="Homo sapiens"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:9606"
ORIGIN

Alignment Scores:
Pred. No.:   0.134   Length: 7396
Score: 176.50   Matches: 170
Percent Similarity: 31.25%   Conservative: 95
Best Local Similarity: 20.05%   Mismatches: 256
Query Match: 5.36%   Indels: 329
DB: Gaps: 34
                                                            6

US-10-736-868-2 (1-643) x AX686226 (1-7396)

Qy 290 oArgLeuLysGluLeu-------------------------------------- 295
Db 4987 GGACCTGAAGGACCGGCGCAACATCGCCTCGGCCAACAAGAACCGGCGACGAGAAGCCAT 5046

Qy 296 -----LeuGlnAsnAlaGluValGlnSerLeuLeuSerTyrGlnArgMetArgAspSe 313
Db 5047 CAAACAGCTGCGAGACGTGCAGAGCCGGGATGAAGGACTGCATGCGCAGCCGAGCTGGATGACAC 5106

Qy 313 rProLeuSerLysArgArgProLeuAlaMetAsnAspGluSerAlaPheArgAl 333
Db 5107 CCGGCGCCTCTCGTGAGGAGATCCTGCCCAGGCCAAGAGAGAAGAAGAGCTGAAGAG 5166

Qy 333 aMetGluAla---------------------------------------ArgAlaLys 339
Db 5167 CATGGAAGCCGAGATGATCCAGTTGCAGGAGGAACTGGCAGCCGCGGAGCGTGCCAAGCG 5226

Qy 339 -- 339
Db 5227 CCAGGGCCCAGCAGGAGCGGGATGAGCTGGCTGACGAGATCGCCACAGCAGCGGCCAAAGG 5286

Qy 340 ----LeuAspGlnLysSerGlnLeuValLeuGlyLeuHisGlyPheGlyGluSe 356
Db 5287 AGCCCTGGCGTTAGAGGAGGAAGCGGGCGTCTGGAGGCCCGATCGCCCAGCTGGAGGAGGA 5346

Qy 356 rAspAspAspGluGluAspAsnIleLeuAspProSerGluAsnSerPheAr 376
Db 5347 GCTGGAGGAGGAGGAGGCAGGCAACACGAGCTGATCAACGAAGGCCCAACCT 5406

Qy 376 gArg---------AlaProLeuArgLeuSerSerGlyPheValGluLysLysLeuLy 391
Db 5407 GCAGATCCGACCAGATCAACGCCGACCTGAACCTGAACCGCGGCGGGCACGCCCAGAAG 5461

Qy 391 sSerAsnAspAspLeuLysSerSerAlaLeuAspArg-----IleLysTyrArgVa 407
Db 5462 ---AACGAGAATGCTCGGCGAACGCTGGAACAAGAAGAGCTTAAGGTCAAGCT 5517

Qy 407 lAspAspValGlu-----LysTyrLeuAla-ProLysPro-------------- 418
Db 5518 GCAGGAGATGAGGCGACTGTCAAGTCCAAGGCCTCCATCACCGGCCCCTGGAGGC 5577

Qy 419 --------MetGluPheAsnProLysPro----------GlnProG 428
Db 5578 CAAGAGATTGCACAGTGCACAGCTGGACAACGAGACCAAGGAGGCGCCAGGCAGCCCTG 5637

Qy 428 lyTyr------------------------------ 429
Db 5638 CAAACAGGTGCCGTCGACCGAGAAGAGCTGAAGGATGTGCTGCTGCAGGTGGATGACGA 5697

Qy 430 ----PheAlaProArgLysIleProThrArgProArgLysMetLeuProLeuLeuG 448
Db 5698 GCGGAGGAACGGCCGAGCAGGTACAAGGACAGCCGACAAGGCATCTACCCGCCTGAGAGCA 5757

Qy 448 lySerAspProLysValGlnGluGluIleArgHisProSer----------- 462
Db 5758 GCTCAAGGCGGCCCAGCGGCCCG---AAGAGGAGGCCCAACGCCCTCCCG 5814

Qy 462 lyTyr------------------------------ 462
Db 5815 CCGGAAACTGCAGCGGCCAGCGTGGAGGACGCCACTGAGACGCCGATGAACCGCGA 5874

Qy 463 ------------------------------ThrG 464
Db 5875 AGTCAGCTCCCTAAAGAACAACAGCTCAGGCGCGGGGACCTGCGTTGTGCTGCCCG 5934

Qy 464 luTrpLys------------------------IleAlaLysGluSerA 472
Db 5935 AATGGCCCGGAAAGGCCCGGGGATGGCTCCGCCGAAGAGGTAGGCCAAAGCCGATGG 5994

Qy 472 rgValLeuThrAsnLeuLysAsnAsnProSerLeuAlaAlaLeuPheMetAspAspLysLysL 492
Db 5995 GGCTGAGGCCCAAACCTGCCGAATAAGCCTCTTCTCTCGCCAGCCTGAGATGGACACAGA 6054

Qy 492 euGluAsnThrLeuLysGlyLysArgGlnMetLeuThrAspArgGlyGlnLysGlyThrArgV 512
Db 6055 CAGACACCACAGCCTCCCCTTCCCAG----ACCCCGCAG :::: 6089

Db 512 aLysThrIleArgArgAlaLeuProArgLeuPheGlyAla----------ProThrA 527
Db 6090 --------CACGGCTCTCCCCCACCTTCTTGGGACTGCGTGTGAACATGGCTCCTCCTG 6138

Qy 527 lAlaLysAlaGluMetIleAspAlaLysLysValPheAlaLysIleGluGlnArgProIleProP 547
Db 6139 CCCTC-----------------------------------CGCCCCGTCCCC 6156

Qy 547 roLeuPhePheGluProLysArgGlyHisThrArgLeuArgTrpThrGlyAlaAsnGluL 567
Db 6157 CATCCCGTTTCCT----------------------------- 6170

Qy 567 ysGlnIleProGlyLeuGlySerArgPheLeuLeuProSerAspPro--------T 584
Db 6171 ------CCAGGTGGTTGTTGAGGAGCATTGGCTTCCTCCTGCTCATCCCCTTCCAGCTC 6222

Qy 584 hrMetProAlaLeuAsnThrAlaPheSerThrGlnGlyArgAlaArgAspAspTrpAspT 604
Db 6223 CCTCCCCGTTCCAGAATCTGATCACCAAAGACAGGGGCCCGGCCCAGGCAGAGAGGACC 6282

Qy 604 hrMetPheLysIleProAsn 610
Db 6283 AGCAGGCTCCTCCAGCCCTCT 6302

RESULT 15
HUMMYONM
LOCUS   HUMMYONM   5122 bp   mRNA   linear   PRI 27-APR-1993
DEFINITION Human nonmuscle myosin heavy chain (NMHC) mRNA, 3' end.
ACCESSION   M31013
VERSION   M31013.1 GI:189035
KEYWORDS   myosin.
SOURCE   Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 5122)
AUTHORS   Saez,C.G., Myers,J.C., Shows,T.B. and Leinwand,L.A.
TITLE   Human nonmuscle myosin heavy chain mRNA: generation of diversity
            through alternative polyadenylylation
JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 87 (3), 1164-1168 (1990)
MEDLINE   90138958
PUBMED   1967836
COMMENT   Original source text: Human macrophage U937, cDNA to mRNA.
            Draft entry and computer-readable sequence for [1] kindly submitted
            by L.A.Leinwand, 21-DEC-1989.
FEATURES   Location/Qualifiers
     source   1..5122
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
     mRNA   <1..5122
                     /product="NMHC mRNA"
     CDS   <2..3745
                     /note="nonmuscle myosin heavy chain (NMHC)"
                     /codon_start=1
                     /protein_id="AAA36349.1"
                     /db_xref="GI:189036"
                     /translation="QEFRQRYEILTPNSIPKGFMDGKQACVLMIKALELDSNLYRIGQ
                     SKVFFRAGVLAHLEEERDLKITDVIIGFQACCRGYLARKAFAKRQQOLTAMKVLQRNC
                     AAYLKLRNWQWWRLFTKVKPLLQVSRQEEMMAKEEELVKREKQLAAENRLMEMETL
                     QSQLMAEKLQLQEOLQAETTELCAEAEELRARLTAKKQELEEICHDLEARVEEEERYQ
                     HLQAEKKMQQONIOELEEOLEEESSAROKLQLEKVTTEAKLKKLBEEQIILEDONCKL
                     AKEKKLLEDRIAEFTTNLITEEEKSKSLAKLKNKHEAMITDLEERLRREEKQRQELEK
                     TRRKLEGDSTDLSDOIAELQAQIAELKMQLAKKEEELQAALARVEEAAQKNMALKKI
                     RELESQISELQEDLESERASRNKAEKQKRDLGEELEALKTELEDTLDSTAAQQELRSK
                     REQEVNILKTLEEEAKTHEAQIQEMRQKHSQAVEELAEQLEQTKRVKANLEKAKQTL
                     ENERGELANEVKVLLQGGRDSEHKRKKVEAQLQELQVKFNEGERVRTELADKVTKLQV
                     ELDNVTGLLSQSDSKSSKLTKDFSALESOLQDTELLQEENRQKLSLSTKLKQVEDEK
                     NSFREQLEEEEEAAKHNLEKQIATLHAQVADMKKKMEDSVGCLETAEEVKRKLQKDLE
                     GLSQRHEEKVAAYDKLEKTKTRLQELDDLLVDLDHQRQSACNLEKKQKKPDQLLAEE"

ORIGIN

```
KTISAKYAEERDRABEAEAREKETKALqSLARALEEAMEQKAELERLNKQFRTEMEDLMS
SKDDVGKSVHELEKSKRALEQQVEEMKTQLELEDELEDEIQATEDAKLRLEVNLQAMKAQF
ERDLQGRDEQSEKKKQLVRQVREMEAELEDERKQRSMAVAARKKLEMDLKDLEAHID
SANKNRDEAIKQLRKLQAQMKDCMRELDDTRASREEILAQAKENEKKLKSMEAEMIQL
QEELAAAERAKRQAQQEBRDELADEIANSSGKGALALEBKRRLEARIAQLEEELEEEQG
NTELINDRLKKANLQIQINADILNLERGHAQKNENARQQLERONKELKVKLQEMEGTV
KSKYKASITALEAKIAQLEEOLDNETKEROAACKQVRRTEKKLKDVLLQVDDERRNAE
QYKDQADKASTRLKQLKRQLEEAEBEAQRANASRRKLQRELEBDATETADAMNREVSSL
KNKLRRGDLPFVVPRRMARKGADGSDEEVDGKADGABAKPAE"
```

Alignment Scores:

Pred. No.:	0.0961	Length:	5122
Score:	175.50	Matches:	170
Percent Similarity:	31.18%	Conservative:	95
Best Local Similarity:	20.00%	Mismatches:	256
Query Match:	5.33%	Indels:	331
		Gaps:	34

US-10-736-868-2 (1-643) x HUMMYONM (1-5122)

```
Qy    12 GlyPheCysIleAlaProLeuSer------AlaGlnSerProSerAsp------     27
            ||||| :: |||| ||
Db  1719 GGCTTCTCAGCCAGTCCGACAGCAGTCCAACCAAGGACTTCTCCGCGTGG        1778

Qy    28 AlaProGly------AlaLeuLeuSerLeuValGlyLysSerHisGlnLeuPro         45
            :: |||||  ||| :::::
Db  1779 AGTCCCAGCTGCAGGACACTCAGGAGCTG-CTGCAGGAGGAGAACCGGCAGAGCTGAGC    1837

Qy    46 LeuAlaProSerMetGluAlaLeuGlu-----                               54
            || ::: ::::::||||
Db  1838 CTGAGCACCAAGCTCAAGCAGGTGGAGGACGAGAAGAATTCCTTCCGGGACGAGAAGCTGGAG  1897

Qy    55 -----LeuMetGlyValGln                                          59
                                  :||
Db  1898 GAGGAGGAGGAGGAGGCCAAGCACAACCTGGAGAAGCAGATGCCACCCTCCATGCCCAG   1957

Qy    60 PheValAspAlaLeuIleLysIleGlyGlnMetGluMetAlaLysPheLysThr        79
            ||||| ::        |||
Db  1958 GTGGGCGAC------ATGAAAAGAAGATGGAGGACAGTGTGGGGTGCCTGGAAACT      2008

Qy    80 GlnLeuGluValLeuGluGlyValHisProAsp------                       90
            |||||  :::::
Db  2009 GCTGAGGAGGTGAAGAGGACCAAGCAGGTTGGAGGGGCCTGAGCCAGCGGCACGAG      2068

Qy    91 -----GlnPheAspLysTyrLysLeuLys------Val                       100
            ::||||||||  |||
Db  2069 GAGAAGGTGGCCGCCTACGACAAGGTGGAGGCCCGAGCCGGCGGGC               2128

Qy   101 AspAspLeuAlaAlaAspAlaValMetGlnGlnAlaGluMetAlaLysLeuGlnLysAsn  120
            |||||  ||  ::::  |||
Db  2129 GACGACCTGCTGGTGGACAAGCTGGAGGACCTGGAGGGCCCGAGCCGCGTGCAACCTGGAAGAGAAG  2188

Qy   121 SerGlyAsnAlaPheIleGluAspMetLeuAsnGlyIleProIleGly------       137
            ||  |||  |
Db  2189 CAGAAGAAG------TTTGACCAGCTCCTGGCCGGGGAGGAGAACCATCTCTCCGCCAAGTAT  2242

Qy   138 ------                                                       138
                  :
Db  2243 GCGAGGAGGACGACGCGCCACCGGGCTGCGAGGCCGAGGAGAAGGACCAAGGCTCTCGTCG  2302

Qy   139 SerIleArgGlyLeuGluAspAlaIleArgThrGlnArgAspMetGluThrAspPro     158
            :::
Db  2303 CTGGACCCGCCCGGGGCTGGAGGAGGCCATGGAGGCCAAAGGCGGCGGGTCAAC--      2359

Qy   159 SerGluGlnIleAlaLysAlaValMetAspLysPheGlnThrGlnIleLeuProGlyLeu  178
            :::||:::
Db  2360 ------AAGCAGTTCCGCACCGAG                                      2377

Qy   179 ValAlaAsnMetIleAlaGlyLysAsnProPhe---LysMetProGlnMetArgLys     197
            ::: :::::::::  |||
Db  2378 ATGGAGGAGACCTTATGAGCTCCAAGGATCGATGTGGCAAGAGTGTCCAAGAGCTGGAGAAG  2437
```

```
Qy   198 AlaGlnAlaAlaProSerSerValPheGlnGlnGlnAlaAlaLeuAlaGlnArgAlaMetLeuGly  217
            |||
Db  2438 TC-CAAGCGGGGCCCT-------AGAGCAGCAGTGGA                         2466

Qy   218 LysAsn------AlaProValAlaGlyGlyArgGlyGluGlnArgMetMetMet       234
            :::
Db  2467 GGAGATGAAGACGCAGCTGGAAGAGCTGAGGCAGCTGCAGGAAGATGCCAA          2526

Qy   235 Asn---ArgValAspGlnArgMetGln-----GlnArgGluLeuGlnGl            248
            |||| :|||||:::|||||
Db  2527 GCTGCGGTTGGAGGTCAACCTGCAGGCCATGAAGGCCCAGTTCAGGACCGAGATGGAACCTGCAGGG  2586

Qy   248 uGluAspGluGluAspAspAspAspLeuGluAspGluIleGlnAlaThrGluAspAla   268
            ::::: ::  ::: :::  |||
Db  2587 CCGGGACGACGAGCAGAGCGAGGAGAAGAAGAAGCAGCTGGTCAGACAGGTGCGGGAGATGGA  2646

Qy   268 pGlyGluProGlnSerGluAlaGluHisGlnArgArgLeuAlaArgArgLeuGluLysSe  288
            |||  :: :::
Db  2647 GGCAGAGCTGGAGGACGAGAGGAAGCAGCGCTCGATGGCAGTGGCCGCGCGGAAGAAGCT  2706

Qy   288 rSerProArgLeuLysGluLeu                                        295
            |||||:::||
Db  2707 GGAGATGGACCTGAAGGACCTGGAGGCCCACATCGACTCGGCCAACAAGAACCGGGACGA  2766

Qy   296 ------LeuGlnAsnAlaGluValGlnSerLeuLeuSerTyrGlnArgMetAr        311
                         |||
Db  2767 AGCCATCAAACAGCTGCGAAGCTGCAGGGCCAGATGAAGGACTGCCATGCGGAGCTGGA   2826

Qy   311 gAspSerProLeuSerLysArgArgProLeuAlaMetAsnAspGluAlaGluSerAlaPh  331
            :::||:::||||
Db  2827 TGACACCCGCGCCTCTCGTGACGAGGATCCTGGCCCCAGGCCAAAGAGAACGAGGAGAGCT  2886

Qy   331 eArgAlaMetGluAla                                              338
                             :||||
Db  2887 GAACAGCATGGAGGCCGAGATGATCCAGTTGCAGGAGGAACTGCCAGCGCGCGAGCGTGC  2946

Qy   338 aLys                                                          339
                 |||
Db  2947 CAAGGCGCCCAGGCCCAGCAGGGGGATGAGCGGGATCTGACGAGATCGCCAACAGCAGCGG  3006

Qy   340 ------LeuAspGlnLysSerGlnLeuValLeuGlyLeuHisGlyPheGl            354
                         |||||  :::::
Db  3007 CAAAGGAGCCCTGGCGTTGAGAGGAGGAAGCGGCGGGTCTGGAGGCCCCATCCCCAGCTGGA  3066

Qy   354 yGluSerAspAspAspGluAspAspGlnAlaGlnLeuGluAsnLeuIleGluAspProSerGluAsnSe  374
            :::::|||
Db  3067 GGAGGAGCTGGAGAGCGGCCAGCAACACGGACCGTGATCAACGAGCTGATCAACGAGAAGGC  3126

Qy   374 rPheArgArg                                                    389
                 : :::
Db  3127 CAACCTGCAGATCGACCAGATCAACGGCCCGACCTGGAGGCGGGCGGGGCACGACGAGAA   3186

Qy   389 sLeuLysSerAsnAspGluLeuLysSerAlaLeuAspArg------IleLysTy       405
            :::::||||
Db  3187 G------AACGAGAATGCTCGGCAGCAGCTGGAACCGGCAGCAGAACAAGGAGCTTAAGGT  3237

Qy   405 rArgValAspAspValGlu------LysTyrLeuAla-ProLysPro---           418
            ::::::                |||
Db  3238 CAAGCTGCAGGAGATGAAGGGGCACTGTCAAGTCAAGGCCTCCAAGTACAAGGCTCCGTCG  3297

Qy   419 ------MetGluPheAsnProLysPro--------------G                   426
                         :::
Db  3298 CGAGGCCAAGATTGCAGAGTTGCAGAGAGCTGACAACAGCGGCGGGTCAAGGC        3357

Qy   426 lnProGlyTyr                                                   429
            |||:||||||:::
Db  3358 AGCCTGCAAACAGGTGCGTCGGACCGAGAGAAAGACTGAAGGATGTGCTGCAGGTGGA   3417

Qy   430 ------PheAlaProArgLysIleProThrArgProArgLysMetLeuProLeuL     446
                         |||       :::
Db  3418 TGACGAGGCGGAGAACCGGCAGAGCTACAGAGCCAGCCGCACGAAGGCCATCTACCCGCCT  3477

Qy   446 euIleGlySerAspProLysProLysValGlnIleLysValGluGluArgHisProSer  462
```

```
Db  3478 GAAGCAGCTCAAGCGGCAGCTGGAGGAGGCCG---AAGAGGAGGCCCAGCGGGGGCCAACGC 3534
                       |||              :::           :::|||    ||||||
Qy   462 ------------------------------------------------------------  462

Db  3535 CTCCGCCGGAAACTGCAGCGCGAGCTGGAGGACGCCACTGAGACGGCCGATGCCATGAA 3594
Qy   462 ------------------------------------------------------------  462

Db  3595 CCGGGAAGTCAGCTCCCTAAAGAACAAGCTCAGGCGGCGGGGGACCTGCCGTTTGTCGTGCC 3654
Qy   463 --ThrGluTrpLys-----------------------------------IleAlaLysG  470
              |||||                                      :::||||||
Db  3655 CCGCCCAATGCCCCGGAAAGGCGCCCGGGATGGCTCCGACCGAAGAGGTAGATGGCAAAGC 3714

Qy   470 luSerArgValLeuThrAsnLeuLysAsnAsnProSerLeuAlaAlaLeuPheMetAspA  490
              :::                  ||||||             ||||
Db  3715 GGATGGGCTGAGGCCAAACCTGCCGAATAAGCCTCTTCTCCTGCAGCCTGAGATGGATG 3774

Qy   490 spLysLeuGluAsnThrLeuLysGlyArgGlnMetLeuThrAspGluGlnLysGlyArgT  510
              :::  :::: |||      |||    |||  :::
Db  3775 GACAGAGACGACCACCAGCCTCCCCTTCCCAG-----ACCCGCAG------------ 3815

Qy   510 hrArgValLysThrIleArgAlaLeuProArgLeuPheGlyAla--------------P  525
              |||                 ||| ||| |||
Db  3816 ----------------CACGCCCTCCCCACCTTCTTGGGACTGCTGTGAACATGCCTC 3858

Qy   525 roThrAlaLysAlaGluMetIleAspAlaLysValPheGlnAspIleGluGluArgProI  545
              |||||                              ||||||: :
Db  3859 CTCCTGCCCTC-----------------------------------CGCCCCG 3876

Qy   545 leProProLeuPhePheGluProLysGlyArgHisThrArgLeuArgTrpThrGlyAlaA  565
              ::|||||||       |||
Db  3877 TCCCCCCATCCCGTTTCCCT---------------------------------- 3896

Qy   565 snGluLysGluIleProGlyLeuGlySerArgPheIleLeuProSerLeuAspPro-----  583
              ||||||||:::                          |||  |||
Db  3897 --------------CCAGGTGTGTTGAGGGCATTTGGCTTCCTCTGCTGCATCCCCTTC 3942

Qy   584 ----ThrMetProAlaLeuAsnThrAlaPheSerThrGlnGlyArgAlaArgAspGluT  602
              ||||||                        :: :
Db  3943 CAGCTCCCTCCCCCTGCTCAGAATCTGATACCAAAAGACACAGGGCCCGGGCCAGGCAGAGA 4002

Qy   602 rpAspThrMetPheLysIleProAsn 610
              |||   ||:::
Db  4003 GCGACCAGCAGGCTCCTCAGCCCTCT 4028
```

Search completed: July 18, 2005, 14:43:23
Job time : 8265 secs

```
                Copyright   GenCore version 5.1.6
                Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on:       July 15, 2005, 13:09:22 ; Search time 924 Seconds
                                          (without alignments)
                                          4119.474 Million cell updates/sec

Title:          US-10-736-868-2
Perfect score:  3291
Sequence:       1 MILFLFLLLGFCIAPLSA.........SKTKRFVGGNGAFDMPALGL 643

Scoring table:  BLOSUM62
                Xgapop  10.0 , Xgapext  0.5
                Ygapop  10.0 , Ygapext  0.5
                Fgapop   6.0 , Fgapext  7.0
                Delop    6.0 , Delext   7.0

Searched:       439206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters:  8780412

Minimum DB seq length:  0
Maximum DB seq length:  2000000000

Post-processing:  Minimum Match 0%
                  Maximum Match 100%
                  Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US10736868/runat_14072005_105509_2268/app_query.fasta_1.839
-DB=N_Geneseq_16Dec04 -QFMT=fastap -SUFFIX=rng -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10736868 @CGN_1_1_470 @runat_14072005_105509_2268 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :    N_Geneseq_16Dec04.*
              1:  geneseqn1980s:*
              2:  geneseqn1990s:*
              3:  geneseqn2000s:*
              4:  geneseqn2001as:*
              5:  geneseqn2001bs:*
              6:  geneseqn2002as:*
              7:  geneseqn2002bs:*
              8:  geneseqn2003as:*
              9:  geneseqn2003bs:*
             10:  geneseqn2003cs:*
             11:  geneseqn2003ds:*
             12:  geneseqn2004as:*
             13:  geneseqn2004bs:*
```

SUMMARIES

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

Result No.	Score	% Match	Query Length	DB	ID	Description
1	181	5.5	3631	4	ABL03905	Abl03905 Drosophil
2	180.5	5.5	7396	12	ADO09868	Ado09868 Human NOV
3	177	5.4	3543	4	AAH62819	^Aah62819 Shrimp wh
c 4	177	5.4	305107	4	AAH62689	Aah62689 Shrimp wh
5	176.5	5.4	7396	6	ABT08505	Abt08505 Human nov

```
 6   174.5  5.3       6432   5   AAS65903   Aa865903 DNA encod
 7   173.5  5.3       7274  13   ADN60282   Adn60282 Human non
 8   173.5  5.2     100990   4   AAK51987   Aak51987 Human pol
 9   171    5.2       7516  12   ADJ94407   Adj94407 Yersinia
10   169.5  5.2       6126  10   ADC13546   Adc13546 Human NOV
11   169.5  5.2       6151   6   ABZ11646   Abz11646 Human pol
12   169.5  5.2       6151  12   ADM44164   Adm44164 Novel hum
13   168    5.1       5439   4   ABL02411   Abl02411 Drosophil
14   168    5.1       5756   4   ABL03904   Abl03904 Drosophil
15   167.5  5.1       6968   4   ABL28617   Abl28617 Drosophil
16   165    5.0       3603  13   ADE49206   Ade49206 Bacterial
17   165    5.0       4755  13   ACN37449   Acn37449 Tumour-as
18   163.5  5.0       2932   9   ADB97572   Adb97572 Rat presy
19   163.5  5.0       6354  10   ADJ56352   Adj56352 Chicken c
20   163    5.0       2485  10   ADF90792   Adf90792 Human hep
21   163    5.0       4041   6   ABQ54663   Abq54663 Human ova
22   162.5  4.9       3257  13   ADS54956   Ads54956 Bacterial
23   162.5  4.9       5883  12   ADQ87119   Adq87119 Human tum
24   162.5  4.9       5883  13   ADQ84854   Adq84854 Human tum
25   162.5  4.9       5883  13   ACN40924   Acn40924 Tumour-as
26   162    4.9       3320  10   ADE85060   Ade85060 Farnesyl
27 c 162    4.9       6607   4   AAH98321   Aah98321 Rabbit ES
28   161.5  4.9       8486   4   AAK52971   Aak52971 Human pol
29   159    4.8       2680   2   AAQ84590   Aaq84590 AMML chro
30   159    4.8       2887   2   AAQ84589   Aaq84589 AMML chro
31   159    4.8       3388   6   ABL61748   Abl61748 Colon ade
32   159    4.8       3388   6   ABL66292   Abl66292 Lung canc
33   159    4.8       3388   6   ABL63718   Abl63718 Breast ca
34   159    4.8       3388   6   ABL64112   Abl64112 Breast ca
35   159    4.8       3388   6   ABL68053   Abl68053 Ovary can
36   159    4.8       3388   6   ABT10826   Abt10826 Human bre
37   159    4.8       4618   3   AAS74130   Aas74130 DNA encod
38   159    4.8       4936   3   AAC61196   Aac61196 Human CDN
39   159    4.8       5574   6   ABK35570   Abk35570 Gene MYH1
40   159    4.8       5835   9   ACA62125   Aca62125 Human smo
41   159    4.8       5937   9   ACA62126   Aca62126 Human smo
42   159    4.8       6861   8   ACF12909   Acf12909 Human cer
43   159    4.8       6861  12   ADK70431   Adk70431 Respirato
44   159    4.8       6861  12   ADN03744   Adn03744 Antipsori
45   159    4.8       6900   8   ACF12910   Acf12910 Human cer
```

ALIGNMENTS

```
RESULT 1
ABL03905
ID   ABL03905 standard; cDNA; 3631 BP.
XX
AC   ABL03905;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster expressed polynucleotide SEQ ID NO 6197.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical; gene; ss.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US009231.
XX
PR   23-MAR-2000; 2000US-0191637P.
PR   11-JUL-2000; 2000US-0614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC, Adams M, Li PWD, Myers EW;
XX
DR   WPI; 2001-656860/75.
```

```
DR   P-PSDB; ABB59802.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signaling and cell-cell
PT   interactions.
XX
PS   Claim 1; SEQ ID NO 6197; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL01840-ABL16175), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
CC   ABB72072). The sequence data for this patent did not form part of the
CC   printed specification, but was obtained in electronic format directly
CC   from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 3631 BP; 1135 A; 941 C; 1080 G; 475 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:              0.000106          Length:        3631
Score:                  181.00            Matches:       165
Percent Similarity:     35.51%            Conservative:  102
Best Local Similarity:  21.94%            Mismatches:    258
Query Match:            5.50%             Indels:        228
DB:                                       Gaps:          37

US-10-736-868-2 (1-643) x ABL03905 (1-3631)

Qy    39 LysSerHisGlnLysProLeu----------------------AlaProSer 49
         |||     :::::||
Db  1031 AAGAGCAACAAGAAGATCCACTGATTGCAGAGGATTCCAAACCAAAAGAAGAGTCCCTAAG 1090

Qy    50 MetGluAlaLeuGluLeuMetGlyValGlnPheValAspAlaLeuIleLysLysGlyGln 69
         |||||||
Db  1091 AAGGAAACAAGCAAAACCGGAGGA--------------------------AAAGGCAAG 1123

Qy    70 MetGluMetAlaLeuAlaGlyAlaPheLysThrGlnLeuGluValLeuGluLysValHisPro 89
         |||||||           ||||||
Db  1124 AAGGAGGAGGAGCAAAACCAGCTGAAAAGAGTAAGAAAGAAAGAAGAA 1171

Qy    90 AspGlnPheAspLysTyrLysLysLeuLysGluValAlaAspAspLeuAlaAlaAspAlaValMet 109
         |||        :::
Db  1172 GACAGCAACAACAAGTCAAAGAAGAAGAAGCAGAGACTCACCTGCAAATAATCAAAAAGAA 1231

Qy   110 GlnGln 111
         |||
Db  1232 CAACAAATTGGAGTCAAGAAAAAGACTTCTGAACCAGAGAGATGCATCGAACTCACACAAA 1291

Qy   112 AlaGluMetAlaLeuLysLeuGlnProLysSerGlyAsnAlaPheIleAspPheMetLeuAsnGly 131
         ::::::
Db  1292 TCGGTGACGGCCAAAAAGGATTCCAAGAAAAGAAGGCATCTTACTTACTCAAGTTAAGTCGGAC 1351

Qy   132 AsnGlyIleProIleGlySerSerIleArgGlyLeu 143
         |||
Db  1352 AACGAATCTCCGGAGCAGCAGAAATCCGATGATCCTCGCACCAGCAGCGAGGCCACAGT 1411

Qy   144 --------GluAspAlaIleArgThrGlnArgMetGlnAsnThrAsp---ProSer 159
         |||                    ::::::
Db  1412 TCCAGCAGGAGCACGAGATGCGGAGATACGGACACCGCCGCGACCAGACCCTCC 1471

Qy   160 GluGlnIleAlaLysAlaValMetAspLysPheGlnThrGlnIleLeuProGlyLeuVal 179
         :::                  |||||||
Db  1472 GCCCAA--------------AAGGAGACTCGCCTGAACAGGAGGAAGAGTCCCGGAA--GTA 1516

Qy   180 AlaAsnMetIleAlaGlyLysAsnPro------PheLysMetProGlnMetArgLys 197
         :::                   :::
Db  1517 ACCGGCGCCCCCAAACAGGGTTCTTACGCCCAAACACGACTCATGGCCAAAAGCAAAAAG 1576

Qy   198 AlaGlnAlaAlaProSerValPheGlnGlnAsnAlaLeuAlaGlnArgGlnAlaMetAlaLeuGly 217
         ||:::   |||||:::
Db  1577 GCTCGCGAGGAGAAG--------GAACAAAGCTGGCGGAGGAGCGCTGCCTCCAAG 1624

Qy   218 LysAsnAlaProValAlaGlyGlyArgGlyGluGlnArgMetMetAsnArgVal 237
         :::
Db  1625 CAACAG--------GATAAGGAGCACCCGAGCAGCAAAAGAAACAGGAGCGC--- 1669

Qy   238 AspGlnArgMetGlnArgIleLeuGln--GluGluAspGluAspAspAspLeu 256
         ||||||         |||:::||
Db  1670 GACGAAAAGGAACAGCAGCGGAAGCTGGAGGAGGATCAGAAGGAGCAGAGAAGATG 1729

Qy   257 GluAspGluAspValProArgArgArgSerSerAspGlyProGlnSerGluAlaGlu 276
         :::    :::   ::   :::::::
Db  1730 GAAAAGGAGGAGGAGGAGGCGCAGCAGCCGGAGGTGGACTCCAGAACGACGAGGAAG 1789

Qy   277 HisGlnArg-------ArgAspLeuAlaArgArgLeuLysSerSerProArgLeuLys 293
         |||
Db  1790 CCCAAGCCCAACCAGGCCAAGGAGGAGTGCAACGCAAAAAGGACGAGGAGCGTCGCAAG 1849

Qy   294 GluLeuLeuGlnAsnAlaGluValGln---- 302
         :::    :::
Db  1850 AAGGAGCAGGAGCGTGAGGAGGCCGACAAAAGAAGAAGAAGCGGCGCCGCCAGTCCTTCTCC 1909

Qy   303 ---------------------------------SerLeuLeu 305
Db  1910 AAGTTCTTCGTGCCCAAGCAGCCCAAGTGCGCGGAGGCGGATCCAACACACATCCTACCTG 1969

Qy   306 SerTyrGlnArgMetArg--AspThrProLeuSerArgArgProLeuAlaMetAsn 324
         :::::::::::
Db  1970 GAGCACGAGCAAAGCAGCTGCCGACAGCTCCAAGGCCAGCAGCCCAAACCGCTG------ 2020

Qy   325 AspGlyGluSerAlaPheGluAlaMetGluAlaAlaArgAlaAlaLysLeu--- 340
         ||||||||||                      :::
Db  2021 ----------GCCTTTCGACCATTCAGATCAAGGAGCACATGCTGCTGCTAGCACCG 2065

Qy   341 ------------------AspGlnLysSerGlnLeu--ValLeuGlyLeuHis 351
         :::::::            |||
Db  2066 ATCGTGCGCAACTCGCTCGGCCCAGGCGCCTCCCAGCTAGATGGACTTTTTCGACAC 2125

Qy   352 GlyPheGlyGluSerAspAspAspGluAspGluGluAspGluAsnLeuIleAspProSer 371
         |||||||
Db  2126 CGGGACGAGGAGGTGACGACGAGGAGGAGGAGGAGGAGGAGCAGGAGCACATCGTC---- 2176

Qy   372 GluAsnSerPheArgArgAlaPro------------LeuArgLeuSer 383
         |||||
Db  2177 ------AGACGAAAGGCCGCCAACATCGGGCTAATCCTCAATCCTATCCTTAGTGAGTTGAGC 2224

Qy   384 SerGlyPheValGluLysLeuLysSerAsnAspGluLeuLysSerAlaLeuAspArgIle 403
         :::::::              |||
Db  2225 AGCGGACGACGGCAAGCCGCCTCAAGATGCAGCCGCGATGTGAAG---CTCCAAAGACGCACC 2281

Qy   404 LysTyrTyrArgValAspAsp------------ValGluLysTyrLeuAlaPro------ 416
         |||        |||
Db  2282 AAGGACGAGGAGGATGATGAAGATGTTCAAGTTGATGATTACCTTTCACCGCGCGGGTCTG 2341

Qy   417 ------------------------------------LysProMetGluPhe 421
         |||
Db  2342 CCCATCGAGTGTGGAGCTGGAGCCCAAGCAGCTGACCCAGCGAATGAAGCCAAGTATCTGCATTTC 2401

Qy   422 AsnProGlnProGlyTyrPhe------------------Ala 431
         :::                               :::
Db  2402 GCCGACAACCCGAGCAGCCACCGTACTACGGCACCTGGGCTAAAAGAGCAGCATCTCC 2461

Qy   432 ProArgLysIleProThrArgProArgLysMetLeuProLeuLeuIleGlySerAspPro 451
         |||||||
Db  2462 GCCCGACGCCGCTGGCCACAGGATAAGGTGCTCTTCGACTACGAAGTGGACTTCCGATTGC 2521

Qy   452 SerThrGluTrpLys 466
         :::
Db  2522 GAGCTGGGAGGAGGAG------GAGCCGGGGAATCGTTGAGCCGTCTAGCGAGGACGAG 2572

Qy   467 IleAlaLysGluArgValLeuThrAsnLeuLysAsnAsnProSerLeuAla------ 484
         :::                  :::
Db  2573 AAAGAGAAGGAGTCCGAGGAGGAGTCAACGAGTACAACGAGTGTACGTGCCGACAC 2632
```

```
Qy   485 AlaLeuPheMetAspAspLysLeuGluAsnThrLeuLysGlyArg------------ 499
             |||:|||       |||
Db  2633 GGACATCTCAGCGACGAGGAACTGCAAAACGACGGCGCGGAATGGAAGACGGCCACACG 2692

Qy   500 ------------------------GlnMetLeuThrAspGlu------GlnLysGlyArg 509
                                 |||:|||        |||:|||
Db  2693 CGGGAGGCGGCAGAAGCGCCAAACTCCAGGTGCTGCAGACAGGAGTTCGGCTCAGGAAATGAAG 2752

Qy   510 ThrArgValLysThrIleArgAlaLeuProArgLeuPheGlyAla------------- 524
                :::       |||:|||       |||
Db  2753 AAGCAGACGACGAAAAGATCAAGGCG------CGCCTGTTCGGTCCCGTTTGGCTCGGACGAG 2806

Qy   525 ------------ProThrAlaLysAspAlaGluMetIleAspAlaAlaLysVal 536
                         |||       |||:|||       |||||||
Db  2807 AACGGCAACAAGTCCGAGCTATTCCCAGCTATCTTTGCCCACACCATCGACATGTACGCC 2866

Qy   537 PheGlnAspIleGluGluGluArgProIleProProLeuPhePheGluGlyProLysGlyGlyArgHis 556
                            :::|||       |||||||
Db  2867 TGTTGGCAGTGGAG------------------CCACTCAGCCTAGAGCCTCC-GCCAGA---- 2907

Qy   557 ThrArgLeuArgTrpThrGlyAlaAsnGluLysGluIleProGlyLeuProGlyLeuGlySerArgPhe 576
             ||||||||||       :::            ||||||
Db  2908 ------------ACCGGAGCGTCAGGATCAGCAACGCGCCAGAGAGCCGGTGCCGGCTGCA 2952

Qy   577 IleLeuProSerLeuAspProThrMetProAlaAlaLeuAsnThrAlaPheSerThrGlnGly 596
                                 :::       :::
Db  2953 GCTGGACGATCG------------TCTCATGCGAGCAGCTGGTGCGCCTTACCCACGGCAATCG 3003

Qy   597 ArgAlaArgAspGluTrpAsp---ThrMetPheLysIleProAsnAsnTrpAsnProGly 615
                            :::|||:||||:|||       ||||||
Db  3004 GAACTCGAAGACGTTCCTGATCAACGAATATTTAGAGTACCTGAA------AACCCAAGC 3057

Qy   616 AspGluValGlyPheLysMetAsnSerLysThrLys 627
             |||       |||:|||       |||:|||
Db  3058 GACCGCCGAGGCCAATCAAACACGCTGCCCTCAA 3093
```

```
RESULT 2
AD009868
ID   AD009868   standard; cDNA; 7396 BP.
XX
AC   AD009868;
XX
DT   01-JUL-2004  (first entry)
XX
DE   Human NOV11 cDNA.
XX
KW   human; ss; gene; NOVX; immunogen; vaccine; cancer; diabetes;
KW   Alzheimer's disease; Parkinson's disease; Huntington's disease; asthma;
KW   allergy; emphysema; bronchitis; arthritis; autoimmune disease;
KW   graft-versus-host disease; arthritis; scleroderma;
KW   systemic lupus erythematosus; bacterial infection; cystic fibrosis;
KW   coronary artery disease; stroke; hypertension; myocardial infarction;
KW   haemophilia; idiopathic thrombocytopenic purpura; hyperlipidaemia;
KW   obesity; cirrhosis; inflammatory bowel disease; Crohn's disease; ulcers;
KW   muscular dystrophy; myasthenia gravis; endometriosis; psoriasis;
KW   alopecia; uveitis; amyotrophic lateral sclerosis; osteoporosis;
KW   osteoarthritis; liver disease; epilepsy; multiple sclerosis; anxiety;
KW   pain; fertility; glomerulonephritis; polycystic kidney disease; SNP;
KW   single nucleotide polymorphism.
XX
OS   Homo sapiens.
XX
FH   Key            Location/Qualifiers
FT   variation      replace(5008,t)
FT                  /tag= a
FT                  /note= "Single nucleotide polymorphism"
FT   variation      replace(5012,g)
FT                  /tag= b
FT                  /note= "Single nucleotide polymorphism"
FT   variation      replace(6808,t)
FT                  /tag= c
FT                  /note= "Single nucleotide polymorphism"
FT   variation      replace(7323,t)
FT                  /tag= d
FT                  /note= "Single nucleotide polymorphism"
XX
PN   US2004018970-A1.
XX
PD   29-JAN-2004.
XX
PF   27-MAR-2002; 2002US-00107782.
XX
PR   19-DEC-2000; 2000US-0256619P.
PR   19-JAN-2001; 2001US-0262959P.
PR   28-FEB-2001; 2001US-0272408P.
PR   28-MAR-2001; 2001US-0279344P.
PR   20-APR-2001; 2001US-0285189P.
PR   26-JUL-2001; 2001US-0308039P.
PR   09-AUG-2001; 2001US-0311266P.
PR   19-DEC-2001; 2001US-0028248.
XX
PA   (SHIM/)  SHIMKETS R A.
PA   (PATT/)  PATTURAJAN M.
PA   (VERN/)  VERNET C A M.
PA   (CASM/)  CASMAN S J.
PA   (MALY/)  MALYANKAR U M.
PA   (SHEN/)  SHENOY S G.
PA   (SPYT/)  SPYTEK K A.
PA   (GANG/)  GANGOLLI E A.
PA   (MILL/)  MILLER C E.
PA   (BOLD/)  BOLDOG F L.
PA   (LILL/)  LI L.
PA   (TAUP/)  TAUPIER R J.
PA   (KEKU/)  KEKUDA R.
PA   (SMIT/)  SMITHSON G.
PA   (ZERH/)  ZERHUSEN B D.
PA   (LIUX/)  LIU X.
PA   (COLM/)  COLMAN S D.
PA   (TCHE/)  TCHERNEV V T.
PA   (SIJJ/)  SI J.
PA   (EDIN/)  EDINGER S R.
PA   (STON/)  STONE D J.
PA   (SCIO/)  SCIORE P.
PA   (MILL/)  MILLET I.
PA   (ROTH/)  ROTHENBERG M E.
XX
PI   Shimkets RA, Patturajan M, Vernet CAM, Casman SJ, Malyankar UM;
PI   Shenoy SG, Spytek KA, Gangolli EA, Miller CE, Boldog FL, Li L;
PI   Taupier RJ, Kekuda R, Smithson G, Zerhusen BD, Liu X, Colman SD;
PI   Tcherney VT, Si J, Edinger SR, Stone DJ, Sciore P, Millet I;
PI   Rothenberg ME;
XX
DR   WPI; 2004-122080/12.
DR   P-PSDB; AD009869, AD009979.
XX
PT   New NOVX polypeptides and nucleic acid molecules, useful for diagnosing,
PT   preventing or treating NOVX-associated disorders e.g. cancer, diabetes,
PT   Alzheimer's disease, stroke, arthritis, hypertension or myocardial
PT   infarction.
XX
PS   Claim 8; SEQ ID NO 35; 240pp; English.
XX
CC   The invention relates to an isolated NOVX polypeptide. A therapeutic,
CC   i.e. the NOVX polypeptide, nucleic acid and antibody, is useful for
CC   manufacturing a medicament for treating a syndrome associated with a
CC   human disease, e.g. a NOVX-associated disorder. The NOVX polypeptides can
CC   be used as immunogens or as vaccines. The NOVX polypeptide, nucleic acid
CC   or antibody is useful for diagnosing, treating or preventing a NOVX-
CC   associated disorder, e.g. cancer, diabetes, Alzheimer's disease,
CC   Parkinson's disease, Huntington's disease, asthma, allergies, emphysema,
CC   bronchitis, autoimmune disease, graft-versus-host disease, arthritis,
CC   scleroderma, systemic lupus erythematosus, bacterial infections, cystic
CC   fibrosis, coronary artery disease, stroke, hypertension, myocardial
CC   infarction, haemophilia, idiopathic thrombocytopenic purpura,
CC   hyperlipidaemia, obesity, cirrhosis, inflammatory bowel disease, Crohn's
```

```
CC  disease, ulcers, muscular dystrophy, myasthenia gravis, endometriosis,
CC  psoriasis, alopecia, uveitis, amyotrophic lateral sclerosis,
CC  osteoporosis, osteoarthritis, liver disease, epilepsy, multiple
CC  sclerosis, anxiety, pain, fertility, glomerulonephritis, or polycystic
CC  kidney disease. The NOVX polypeptides and nucleic acid molecules are
CC  useful for determining the presence of or predisposition to a disease
CC  associated with altered levels of the NOVX polypeptide or the nucleic
CC  acid molecule, or for screening for molecules that inhibit or enhance
CC  NOVX activity or function. The nucleic acids may be used as hybridisation
CC  probes, in chromosome mapping, tissue typing, preventive medicine, or
CC  pharmacogenomics. The present sequence represents cDNA encoding a human
CC  NOVX protein.
XX
SQ  Sequence 7396 BP; 1878 A; 2016 C; 2290 G; 1212 T; 0 U; 0 Other;
```

```
Alignment Scores:
Pred. No.:             0.000281        Length:        7396
Score:                 180.50          Matches:       162
Percent Similarity:    30.11%          Conservative:  96
Best Local Similarity: 18.90%          Mismatches:    253
Query Match:           5.48%           Indels:        347
DB:                                    Gaps:          32
```

US-10-736-868-2 (1-643) x AD009868 (1-7396)

```
Qy   12 GlyPheCysIleAlaProLeuSer------AlaGlnSerProSerThrSerAsp------ 27
Db 3999 GGCTTCTCAGCCAGTCCGACAGCAAGTCCAGCAAGGACTTCTCCGCGCTGG 4058

Qy   28 AlaProGly-----AlaLeuLeuSerLeuValGlyLysValGlySerHisGlnLysLeuPro 45
Db 4059 AGTCCCAGCTGCGAGGAACACTCAGGAGCTG-CTGCAGGAGGAACCGGCAGGAAGCTGAGC 4117

Qy   46 LeuAlaProSerMetGluGluAlaLeuGlu------ 54
Db 4118 CTGAGCACCAAGCTCAGCAGTGGAGGACCAGGAGAATTCCTTCCGGAGCCAGCTGGAG 4177

Qy   55 ------LeuMetGlyValGlnPheVal 61
Db 4178 GAGGAGGAGGCCAAGCACAACCTGGAAGAAGCAGATGCGCCACCTCCATGCCCAGGTGGCC 4237

Qy   62 AspAlaLeuIleLeuLysLysGlyLysGlnMetGluAlaLysAlaLysPheLysThrGlnLeu 81
Db 4238 GAC------ATGAAAAGAAGATGGAGACAGTGGAGGGGGTGCCTGGAAACTGCTGAG 4288

Qy   82 GluValLeuGluLysLysValHisProAsp-- 90
Db 4289 GAGGTGAAGAGGAAGCTCCAGAAGGACCTGGAGGGGCCTGAGCCAGCGGCACGGAGGAGAAG 4348

Qy   91 ------GlnPheAspLysTyrLysLeuLys------ValAspAsp 102
Db 4349 GTGGCCGCCTACGACAAGTGGAAGAAGACCGCGGCGTGCAGCAGGAGCTGGACGAG 4408

Qy  103 LeuAlaLeuAspAlaValMetGlnGlnLeuAlaGluMetGlnLysLysLeuLysGlnProLysSerGly 122
Db 4409 CTGCTGGTGGACCAGGTGGAGGCGCCGGGCCGAGAGGAGAAGCAG 4468

Qy  123 AsnAlaPheIleAspMetLeuAsnGlyAsnGlyIleProIleGly------ 137
Db 4469 AAG-----TTTGACCAGCTCTGGCCAGGAAGACCATTCTCTGCCAAGTATGCAGAG 4522

Qy  138 ------SerSerIle 140
Db 4523 GAGGCGGCGGCTGAGGCGCCGGAGGCGCCGAGGAGAAGACCAAGCCTCTGTCGCTGGCC 4582

Qy  141 ArgGlyGluAspAlaIleAlaArgThrGlnLysArgAspMetGluAsnThrAspProSerGlu 160
Db 4583 CGGGGCCCTGAGGAGAAGCCATGGAGAAGGCGGAGCGGGAGCTGGAGCGCGGCGCTCAAC-- 4633

Qy  161 GlnIleAlaLysAlaValMetGluLysPheGlnThrGlnIleLeuProGlyLeuValAla 180
Db 4634 AAGCAGTTCCCACCGAGGATGGAGAGGACCTTATGAGCTCC 4672
```

```
Qy  181 AsnMetIleAlaGlyLysAsnProPheLysMetProGlnMetArgLysAlaGlnAla 200
Db 4673 AAGGATGATGTGGCCAAGAGTGTCGAGAGTGGAGAGTCCAAGCCG------------ 4720

Qy  201 AlaProSerSerValPheGlnGlnLeuAlaGluLeuArgAlaMetLeuGlyLysAsnAla 220
Db 4721 ------------------GCCCTAGAGCAGCAGGTG-- 4738

Qy  221 ProValAlaGlyGlyArgGlyGluGlnArgMetMetAsnArgValAspGlnArg 240
Db 4739 ------GAGGAGATGAAGAGCCAGCTGGAAGAGCTGGAGCACGAG 4777

Qy  241 MetGln 242
Db 4778 CTGCAGGACCAGGTGCGGAGGAAGATGCCACGACCTGCGGTTGGAGTCAACCTGCACCATGAAGGCC 4837

Qy  243 ------GlnArgGluLeuGlnGluAspLeuAspAspAspLeuGluAspLeuGluAsp 260
Db 4838 CAGTTCGAGCGGGACCTGCAGGCCCGGAGCAGCAGCAGCAGGAGGAGAAGAAGCAGCTG 4897

Qy  261 ValProArgArgSerAspSerArgProGlnSerGluValAlaGluHisGlnHisGlnArgArg 280
Db 4898 GTCAGACCAGGTGCGGAGACAGCTCAGCAGATGGAGCAGAGGAGGACGAGGAGCAGCGCTCGATG 4957

Qy  281 AspLeuAlaArgArgLeuLysSerSerProArgLeuGlyLeuGlu------ 295
Db 4958 GCAGTGGCGCGGCGGAGAAGTGGAGATGGACCTGAAGGACCTGGAGGCCGCACATCGAC 5017

Qy  296 ------LeuGlnAsnAlaGluValGlnSer 303
Db 5018 TCGGCCAACAAGAACCGGACGAAGCCATCAAACAGCTGCGGAAGCTGCAGGCCCAGATG 5077

Qy  304 LeuLeuSerTyrGlnArgMetArgAspSerProLeuSerArgArgProLeuAlaMet 323
Db 5078 ATGGACTGCATGCCAGCTGGATGACACCCGCGCCTTCGTGAGGATCCTTGGGCCCAG 5137

Qy  324 AsnAspGluSerAlaAlaPheArgAlaMetGluAla 336
Db 5138 GCCAAAGAGAACGAGAAGAGCATGGAGGGCCGAGATGATCCAGTTGCAGGAG 5197

Qy  337 ------ArgAlaLys 339
Db 5198 GAACTGCAGCCGCCAAGGCGTGCCAGGCCCAGCAGGAGCGGGATGAGCTGGCT 5257

Qy  340 ------LeuAspGlnLysSerGlnLeu 346
Db 5258 GACGAGATCGGCAACAGCAGCGGGCAAAGGAGGCCCTGGCGTTAGAGAGGAAGGCGGTCTG 5317

Qy  347 ValLeuGlyLeuHisGlyPheGlyGluSerArgAspAspArgGluGlyGluAspGluAsn 366
Db 5318 GAGGCCCCATGCCGTCGGAGGAGGAGGCCGCAACACCGGACGAGAGCTG 5377

Qy  367 LeuIleAspProSerGluAsnSer------ 381
Db 5378 ATCAACGACCGGCTGAAGAAGGTCAAGCTGCAGATCGACCATCAACGCCGACCTGAC 5437

Qy  382 LeuSerSerGlyPheValGlyLysSerSerAsnAspGluLeuLysSerAlaLeuAspAsp 401
Db 5438 CTGGAGCGCGGGCGCACCGCCAGAG------AACGAGAATGCTCGGCAGCAGCTGGAA 5488

Qy  402 Arg------IleLysTyrArgValAspValGlu------Lys 412
Db 5489 CGCCAGAACACCGGCTTAAGGTCAAGCTGCAGGAGATGGAGGGCACTGTGCAAGTCCAAG 5548

Qy  413 TyrLeuAla-ProLysPro------MetGluPheAsp 422
Db 5549 TACAAGGCCTCCATCACCGCCCTCGAGGCCAAGATTGCACAGTCGCAGCTGGAC 5608

Qy  422 nProLysPro------GlnProGlyTyr 429
Db 5609 AACGAGACCAAGGAGGCGCCAGGCAGCTGCAAACAGGTGCGTCGGACCGAGAAGAAGCTG 5668

Qy  430 ------PheAlaProArgLysIleProThrAr 438
```

```
XX
PR   24-NOV-1999;  99CN-00124717.
XX
PA   (PENY-) PE CORP NY.
PA   (THIR-) THIRD INST OCEANOGRAPHY STATE OCEANI C A.
PA   (SINO-) SINOGENOMAX CO LTD.
XX
PI   Xu X,  Yang F,  He J,  Pham L,  He M,  Ye Y,  Shen Y,  Kodira C;
XX
DR   WPI; 2001-355877/37.
DR   P-PSDB; AAG85039.
XX
PT   Primary nucleotide sequence of the shrimp white spot Bacilliform virus
PT   (WSBV), useful for producing viral polypeptides that can be used to
PT   screen for agents that are useful for treating WSBV infection.
XX
PS   Claim 4; Fig 3; 626pp; English.
XX
CC   The invention provides the primary nucleotide sequence of the WSBV genome
CC   (AAH62689), predicted transcript sequences (AAH62689-AAH62839) and
CC   encoded proteins (AAG84910-AAG85051) and oligonucleotide sequences
CC   (AAH62840-63160) suitable for use as primers or probes. The nucleic acid
CC   molecules and proteins of the invention are useful for diagnosis and
CC   monitoring viral infection, in screens for antiviral agents and for
CC   monitoring viral gene expression or activity during a treatment regimen.
CC   The nucleic acid molecules are also useful as antisense constructs to
CC   control viral gene expression in infected cells and tissues and to create
CC   transgenic viral resistant shrimp. (Updated on 06-AUG-2003 to correct OS
CC   field.)
XX
SQ   Sequence 3543 BP; 1288 A; 625 C; 823 G; 807 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:                     0.000218        Length:               3543
Score:                       177.00            Matches:                99
Percent Similarity:           35.39%           Conservative:           67
Best Local Similarity:        21.11%           Mismatches:            165
Query Match:                   5.38%           Indels:                138
DB:                           4                Gaps:                   20

US-10-736-868-2 (1-643) x AAH62819 (1-3543)
```

```
Qy   176  ProGlyLeuValAlaAlaAsnMetIleAlaGlyLysAsnProPheLysMetProGlnMet 195
              |||  |||  |||      |||      :::::  |||||
Db   427  CCTGCGTTATTGGCGTAGCGGCAACCAACAAGAAGATGCCGACAATTATACCTGAAAACCAG 486

Qy   196  ArgLysAla-------------GlnAlaAlaProSerSerValPheGlnGlnAlaLeuAlaGln 212
              |||:|||                :::  |||||:|||
Db   487  TGGAAAGTATAGTGAACACCCCCGCCCAGTAGGGCCAAATAGGCAAGTTCTTCCCATG 546

Qy   213  ArgAlaMetLeuGlyLysAsnAlaProValAlaGlyGly---ArgGlyGluGlnArg 231
              |||                                |||||||||
Db   547  TTGAACTTTTTACTTGAAAATGTGAACGGCCATGGCGGCGGATCAGCAGGTGAAGAACAGAAG 606

Qy   232  MetMetMetAsnArgValArgMetGlnArgMetLeuGlnLeuGlnGluGluAspGlu 251
                                       |||:||  |||        |||:|||||
Db   607  ------------AACAAAGAAGATGATAAC---CAACAAATAGAACCCTAGAAGAAGAGAG 654

Qy   252  AspAspAspLeuGluAspGluAspValProArgArgArgSerAspGlyGluPro 271
              |||:|||  |||:||||                         |||||
Db   655  GATGAAGAGGAGGAACAGAGGAG--------------GAAGAGGAAGAGGAAGAGAGAA 702

Qy   272  GlnSerGluAlaGluHisGlnArgArgMetGlnArgArgAspArgArgLeuLysArgLeuAsArgSerSerProArg 291
              |||  |||  |||      ||||    :::  |||      :::
Db   703  CAGGAGGAGGAAGAGGAAGAGGAAGAAAGAAGAACCTATAGAACAAGAAAAAATGAACCCGGAGAAG 762

Qy   292  LeuLysGluLeuGluLeuGlnAsnAlaGluLeuValGlnLeuSerLeuLeuSerTyrGlnArgMetArg 311
              |||:|||  :::::|||||                           |||||
Db   763  GATGAAGATGCAATAGAAAATGAAAAGCGTGCACTCT----------CATGAGTTGAA 810

Qy   312  AspSerProLeuSerLysTyrArgArgProLeuAlaMetAsnAspGluAlaAspGluSerAlaPhe 331
              |||:||:::                            |||:||
Db   811  TCCAGCCCATGAGTGAA----------------GGAGGTAATGATGATGAAGATGGATTATTC 858
```

```
Db   5669  AAGGATGTGTCTGCTGCAGGTGGATGGACGCGACGAGCGGAGGAACGCCGAGCAGTACAAGGACCAG 5728
                                                            |||||
Qy   438  gProArgLysMetLeuProLeuLeuIleSerGluAspProLysValGlnGluSerAspLysGluIleLeuAr 458
               |||  :::  ||||||       |||            ::
Db   5729  GCCGACAAGCAATCACCCGCCTCAAGCAGCTTCAAGCGGCAGCTGGAGGAGGCCG---AA 5785

Qy   458  gArgHisProSer------------------------------ 462
               :||  ||||||
Db   5786  GAGGAGGCCCAGCGGCGGCCAACGCCTCCCGCCGGAAACTGCAGCGCGAGCTGGAGGACGCC 5845

Qy   462  ------------------------------------------ 462
Db   5846  ACTGAGACGGCCGATGCCATGAACCGGCAGGTCAGCTCCCTAAAGAACAAGCTCAGGCGC 5905

Qy   463  ThrGluTrpLys----------------------------- 466
               |||||||
Db   5906  GGGGACCTGCCGTTTGTCGTGCCCCGGCAATGGCCCGGAAAGGCGCCGGGGATGGCTCC 5965

Qy   467  ------------IleAlaLysGluSerArgValLeuThrAsnLeuLysAsnAsnProSe 482
                          :::|||||||       ::        |||      :::
Db   5966  GACGAAGAGGTAGATGCCAAAGCCGATGGGGCTGAGGCCCAAACCTGCCGAATAAGCCTCT 6025

Qy   482  rLeuAlaAlaLeuPheMetAspAspAlaAspLysLeuLeuLysAsnThrLeuLysGlyArgGlnMetLe 502
               |||       |||  :::  :::  |||      |||
Db   6026  TCTCCTGCAGCCTGAGATGGATGACAGACAGAGCACCACCAGCCTCCCCTTCCCAG------ 6080

Qy   502  uThrAspGluGlnLysGlyArgThrArgValLysThrIleArgAlaLeuProArgLeuPh 522
               |||  :::                                |||  |||
Db   6081  -ACCCCGCAG-----------------------------CACGCCCTCTCCCCACCTTCT 6109

Qy   522  eGlyAla-------------ProThrAlaLysAlaGluMetIleAlaAspAlaLysValPh 537
               |||||                |||      |||
Db   6110  TGGGGACTGCTGTGAACATGCCTCCTCCTGCCCTC--------------------------- 6143

Qy   537  eGlnAspIleGluGluGlnProIlePheProLeuPhePheGluProLysGlyArgHisTh 557
               |||  |||  |||                      ||||||:|||||
Db   6144  ---------------------------CGCCCGTCCCCCATCCGTTTCCT------------- 6170

Qy   557  rArgLeuArgTrpThrGlyArgAlaAsnGluLysGlyIleLeuProGlyLeuGlySerArgPheIl 577
                                               |||||||:::
Db   6171  ---------------------------CCAGGTGTTGTTGAGGGCATTTG 6193

Qy   577  eLeuProSerLeuAspPro---------ThrMetProAlaLeuAsnThrAlaPheSerThr 594
               |||||                      |||  :::  |||
Db   6194  GCTTCCTCTGCTGCATCCCCTTCCAGCTCCCCCTGCTCCCCTGTCAGAATCTGATACCCAAAGAG 6253

Qy   594  rGlnGlyArgAlaArgAspAspGluTrpGluTrpAspThrAspThrMetPheLysIleProAsn 610
                                |||  |||                       |||||
Db   6254  ACAGGGCGCCGGGCCAGGCAGAGAGCGACCAGCAGGCTCCTCAGCCCTCT 6302
```

```
RESULT 3
AAH62819
ID   AAH62819 standard; DNA; 3543 BP.
XX
AC   AAH62819;
XX
DT   06-AUG-2003 (revised)
DT   11-SEP-2001 (first entry)
XX
DE   Shrimp white spot Bacilliform virus (WSBV) gene 130.
XX
KW   Shrimp white spot Bacilliform virus; WSBV; diagnosis; viral infection;
KW   antiviral agent; gene expression; antisense construct;
KW   transgenic viral resistant shrimp; ds.
XX
OS   Shrimp white spot syndrome virus.
XX
PN   WO200138351-A2.
XX
PD   31-MAY-2001.
XX
PF   08-NOV-2000; 2000WO-US028888.
```

```
XX
DE   Shrimp white spot Bacilliform virus (WSBV) genomic sequence.
XX
KW   Shrimp white spot Bacilliform virus; WSBV; diagnosis; viral infection;
KW   antiviral agent; gene expression; antisense construct;
KW   transgenic viral resistant shrimp; ds.
XX
OS   Shrimp white spot syndrome virus.
XX
PN   WO200138351-A2.
XX
PD   31-MAY-2001.
XX
PR   08-NOV-2000; 2000WO-US028888.
PR   24-NOV-1999; 99CN-00124717.
XX
PA   (PENY-) PE CORP NY.
PA   (THIR-) THIRD INST OCEANOGRAPHY STATE OCEANI C A.
PA   (SINO-) SINOGENOMAX CO LTD.
XX
PI   Xu X, Yang F, He J, Pham L, He M, Ye Y, Shen Y, Kodira C;
XX
DR   WPI; 2001-355877/37.
XX
PT   Primary nucleotide sequence of the shrimp white spot Bacilliform virus
PT   (WSBV), useful for producing viral polypeptides that can be used to
PT   screen for agents that are useful for treating WSBV infection.
XX
PS   Disclosure; Fig 2; 626pp; English.
XX
CC   The invention provides the primary nucleotide sequence of the WSBV genome
CC   (AAH62689), predicted transcript sequences (AAH62689-AAH62839) and
CC   encoded proteins (AAG84910-AAG85051) and oligonucleotide sequences
CC   (AAH62840-63160) suitable for use as primers or probes. The nucleic acid
CC   molecules and proteins of the invention are useful for diagnosis and
CC   monitoring viral infection, in screens for antiviral agents and for
CC   monitoring viral gene expression or activity during a treatment regimen.
CC   The nucleic acid molecules are also useful as antisense constructs to
CC   control viral gene expression in infected cells and tissues and to create
CC   transgenic viral resistant shrimp. (Updated on 06-AUG-2003 to correct OS
CC   field.)
XX
SQ   Sequence 305107 BP; 92042 A; 62482 C; 62635 G; 87948 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:                0.0536       Length:           305107
Score:                    177.00       Matches:              99
Percent Similarity:       35.39%       Conservative:         67
Best Local Similarity:    21.11%       Mismatches:          165
Query Match:              5.38%        Indels:              138
DB:                            4       Gaps:                 20

US-10-736-868-2 (1-643) x AAH62689 (1-305107)
```

```
Qy  332 ArgAlaMetGluAlaAlaArgAlaLysLeuAspGlnLysSerGlnLeuValLeuGlyLeuHis 351
            :::        |||        :::
Db  859 TTCTCCTCAATAGCAGGTGGCGGCGGCCAATGATAATGAAGAGGAT----------------- 900

Qy  352 GlyPheGlyGluSerAspAspAspGluGluAspGluGluAsnLeuAsnGluIleAspProSer 371
            ::::  |||    ::::        |||    :::
Db  901 ------GAAGAAGGATGAAGAGAAGAGGGAGGAGGAGGAGGAACCTGCA 951

Qy  372 GluAsnSerPheArgArgAlaProLeuArgLeuSer--SerGlyPheValGluLysLeu 390
            :::        :::
Db  952 CAAAAATCTGAAGAGCATGTAGAAACTAAGAGTCTGTTCAATCTCACACGGAATATATT 1011

Qy  391 LysSerAsnAspGluLeuLysSerAlaLeuLeuAspArgIleLysTyrArgValAspAspVal 410
            :::     :::            :::            |||   ::::::  |||
Db 1012 GAAGAAGAGGAAGAAGAATATGAAGAAATGAAGACTCCGGCACCACTCGAGGATGA 1071

Qy  411 GluLysTyrLeuAlaProLysProMetGluPheAsnProLysProGlnProGlyTyrPhe 430
            :::     ::::||:
Db 1072 GAA--------ATAAGTACTATGCATCAATTTAACAATGCGCCCAGGGGTTCGAAGATCT 1122

Qy  431 AlaProArgLysIle 435
            ::::
Db 1123 CCCCCTCCAGATATACAAGAGTGCGAGGATGCAGTTGTGTTCCCTCCAATCATGAAAGAG 1182

Qy  436 ----------------------ProThrArgProArgLysMetLeuPro 444
                                  |||     |||||
Db 1183 ACAGATATTCTCCCCCAGATAAAGGAACCTTCACCAAAGCCACCTAGAAGTGTTTTCTATT 1242

Qy  445 LeuLeuIleGlySerAsp----------------------ProLysVal 453
            |||                                   |||||
Db 1243 CTTGGTAGTGGAGGTGAGGAACAATATGACCAACTAAACGATATAGCACCACCGCCGTGTT 1302

Qy  454 GlnGluGluIleArgArgHisProSerThrGluTrpLysIleAlaLysSerArg---------- 472
            :::           |||   :::|||||||
Db 1303 CCATTCATCGTGACTTTTCCTCCAGACAATGAA------ATGGGAGAAGAAAGTAGAGAC 1356

Qy  472 -------------------------------------------- 472
Db 1357 ATCATGGACCCAAGATCGATGCTGATGCCTCCTCCTCCCACCTCCACCACCACCT 1416

Qy  472 -------------------------------------------- 472
Db 1417 CATCAACCTCCTCAACTCAAACCTACCAATATTCTTCTTCCCTCCACCTCCACCT 1476

Qy  473 ----------------ValLeuThrAsnLeuLysAsnAsnProSerLeuAlaAlaLeuPheMet 488
                            :::   ::|||||
Db 1477 ACAAATCAGTCCTTATTTTCAAACAACAACAATCCATCCTTTTTGAGTACAGTGGTT 1536

Qy  489 AspAspGlyLeuGluAsnThrLeuLysGlyArgGlnMetLeuThrAspGluGlnLysLysGly 508
            :::        |||||  :::|||
Db 1537 ---GGAAAGGTTAACACACATTAGGTGGGAAGAGGCCGAAGAAAGATTACACAGACT 1593

Qy  509 -------------ArgThrArgValLysThrArgIleArgAlaLeuProArgLeuPhe 522
                         :::|||
Db 1594 ATGGAGTCTATAAATCTTAAAAACAAGGGTAAAAACTTCTT----------------- 1632

Qy  523 GlyAlaProThrAlaLys-----------AlaGluMetIleAspAlaLysValPheGln 538
            :::        :::||:
Db 1633 ---CTAGAAACAACAAGAACCTTCAATGCTCTGAACTCGGTGAAAGTT---GTGTTCCAA 1686

Qy  539 AspIleGluGluArgProIleProPro 547
            |||::        :::
Db 1687 GAT---CCAGAAAACCCAGTAAAACCA 1710

RESULT 4
AAH62689/C
ID   AAH62689 standard; DNA; 305107 BP.
XX
AC   AAH62689;
XX
DT   06-AUG-2003 (revised)
DT   11-SEP-2001 (first entry)
```

```
Qy  176 ProGlyLeuValAlaAlaAsnMetIleAlaGlyLysAsnProPheLysMetProGlnGlnMet 195
            |||      :::|||                       ::::::
Db 183391 CCTGCGGTTATTGGCGCTAGCGCCACACAAGAGATGCGACAATTATACCTGAAAACCAG 183332

Qy  196 ArgLysAla----------GlnAlaAlaProSerValPheGlnAlaGlnAlaLeuAlaAlaGln 212
            :::|||              :::|||
Db 183331 TGGAAAGTATAGTGACACACCCCTCGCCAGTAGGGCCAAATAGGCAAGTTCTTCCCATG 183272

Qy  213 ArgAlaMetLeuGlyLysAsnAlaProValAlaGlyGly--ArgGlyGluGluGlnArg 231
            |||      :::|||||||||||||                 |||||
Db 183271 TTGGAACTTTTTACTTGAAATGTGAAGCCATGGGGCCGGATCAGGTGAAGAACAGAAG 183212

Qy  232 MetMetMetAsnArgValAspAspArgMetGlnLysArgMetGlnArgGluGluGlnGluGlu 251
            |||::        |||::        |||::|||||
Db 183211 ------ACAAAGAAGATGATAAC---CAACAAATAGAACCCGTAGAAGAGGAAGAG 183164

Qy  252 AspAspAspAspLeuGluAspGluAspProArgArgArgSerArgAspGlyGluGluPro 271
            |||::::::::                   :::            |||
Db ...
```

Qy   539 AspIleGluGlnArgProIleProPro 547

Db   182131 GAT---CCAGAAAACCAGTAAAACCA 182108

RESULT 5
ABT08505
ID   ABT08505 standard; cDNA; 7396 BP.
XX
AC   ABT08505;
XX
DT   28-NOV-2002 (first entry)
XX
DE   Human novel protein NOV11 coding sequence SEQ ID NO: 35.
XX
KW   Human; NOVX; single nucleotide polymorphism; SNP; anti-HIV; cytostatic;
KW   antiarteriosclerotic; antidiabetic; antiasthmatic; antiinflammatory;
KW   haemostatic; hypotensive; neuroprotective; anorectic; nootropic;
KW   antidepressant; immunosuppressive; antibacterial; antiparasitic;
KW   virucide; tranquilizer; anticonvulsant; osteopathic; analgesic;
KW   antiparkinsonian; dermatological; antiinfertility; cerebroprotective;
KW   antiaddictive; gene; ss.
XX
OS   Homo sapiens.
XX
FH   Key   Location/Qualifiers
FT   variation   replace(5008,T)
FT   /*tag= a
FT   variation   replace(5012,G)
FT   /*tag= b
FT   variation   replace(6808,T)
FT   /*tag= c
FT   variation   replace(7323,T)
FT   /*tag= d
XX
PN   WO200259315-A2.
XX
PD   01-AUG-2002.
XX
PF   19-DEC-2001; 2001WO-US050076.
XX
PR   19-DEC-2000; 2000US-0256619P.
PR   19-JAN-2001; 2001US-0262959P.
PR   28-FEB-2001; 2001US-0272408P.
PR   20-APR-2001; 2001US-0285189P.
PR   26-JUL-2001; 2001US-0308039P.
PR   09-AUG-2001; 2001US-0311266P.
XX
PA   (CURA-) CURAGEN CORP.
XX
PI   Shimkets RA, Patturajan M, Vernet CAM, Casman SJ, Malyankar U;
PI   Shenoy S, Spytek KA, Gangolli E, Miller C, Boldog F, Li L;
PI   Taupier RJ, Kekuda R, Smithson G, Zerhusen BD, Liu X, Colman SD;
PI   Tcherney V, Si J, Edinger S, Stone D, Sciore P, Millet I;
PI   Rothenberg M;
XX
DR   WPI; 2002-666903/71.
DR   P-PSDB; ABJ10604.
XX
PT   New isolated NOVX polypeptides and polynucleotides, useful for
PT   preventing, diagnosing or treating NOVX-associated disorders e.g.
PT   diabetes, Crohn's disease, atherosclerosis, cancer, Huntington's disease
PT   or Alzheimer's disease.
XX
PS   Claim 8; Page 135-137; 363pp; English.
XX
CC   The present invention provides the protein and coding sequences of
CC   several novel human proteins, designated NOVX. These can be used in the
CC   treatment of diseases such as cancers, Hodgkin's disease, Von Hippel-
CC   Lindau syndrome, Alzheimer's disease, stroke, tuberous sclerosis,
CC   hypercalcaemia, Parkinson's disease, Huntington's disease, cerebral
CC   palsy, epilepsy, Lesch-Nyhan syndrome, multiple sclerosis, ataxia
CC   telangiectasia, leukodystrophies, addiction, anxiety, depression, pain,

Db   183163 GATGAAGAGAGAGAGGAAGAACAGGAGGAG-------------GAAGAGGAAGAGGAAGAAGAA 183116

Qy   272 GlnSerGluAlaGluHisGlnArgAspLeuLeuArgArgLeuLeuLysSerSerProArg 291
                ||| ||| ||| :::   :::
Db   183115 CAGGAGGAGGAGGAGGAAGGAGGAAGAAAAGAACCTATAGAACAAGAAAAAATGAACCCGAGAAG 183056

Qy   292 LeuLysGluLeuGlnAsnAlaGluValGlnSerLeuLeuSerTyrGlnArgMetArg 311
                ::::|| |||:::|||:: ||
Db   183055 GATGAAGATGCAATAGAAATGAAAGCGTGCACTCT---------CATAGAGTTGAA 183008

Qy   312 AspSerProLeuSerLysArgArgProLeuAlaMetAsnAspGluSerAlaPhe 331
                |||::||::|||||||:||
Db   183007 TCCAGCCCTATGAGTGAA---------GGAGGTAATGATGATGGAATGGATTATTTC 182960

Qy   332 ArgAlaMetGluAlaArgAlaLysArgLeuLeuAspAspSerGlnLeuValLeuGlyLeuHis 351
                :::
Db   182959 TTCTTCAATAGCAGGTGGCGCGAATGAATAATGAAGAGGAT------------------ 182918

Qy   352 GlyPheGlyGluSerArgAspAspPheGluGlyLeuAsnLeuIleAspProSer 371
                :::||||:::||||||| |||:::
Db   182917 -------GAAGAGAGGATGAAGAAGAGGAGGAGGAGGAGGAGGAGGAACCTGCA 182867

Qy   372 GluAsnSerPheArgArgAlaProLeuArgLeuSer---SerGlyPheValGluLysLeu 390
                ::: |||
Db   182866 CAAAAATCTGAAGAGCATGTAGAAACTAAAGAGTCTGTTCAATCTCACACGGAATATATT 182807

Qy   391 LysSerAsnAspGluLeuLysSerAlaLeuAspArgIleLeuTyrArgValAspAspVal 410
                :::||| ||| :::::|||
Db   182806 GAAGAAGAAGAATATGAAGAGATGAAATCTCGGCACACTCGAGGATGAA 182747

Qy   411 GluLysTyrLeuAlaProLysProMetGluPheAsnProLysProGlnProGlyTyrPhe 430
                :::||||||:||
Db   182746 GAA------ATAAGTACTATGCATCAATTTAACAATGCGCCCAGGGTTCGAAGATCT 182696

Qy   431 AlaProArgLysIle 435
                |||
Db   182695 CCCCCTCCAGATATACAAGAGTGCGAGGATGCAGTTGTGTCCCTCCAATCATGAAAGAG 182636

Qy   436 ProThrProArgLysMetLeuPro 444
                :::||||||
Db   182635 ACAGATATTCTCCCCCAGATAAAGAACCTTCACCAAAAGCACCTAGAATGTTTTCTATT 182576

Qy   445 LeuLeuIleGlySerAsp 453
                ||| ||
Db   182575 CTTGGTAGTGGAGGTGAGGAACACATATGACCACTAAACGATATAGCACCGCCTGTT 182516

Qy   454 GlnGluGluIleArgArgHisProSerThrGluTrpLysIleAlaLysGluSerArg--- 472
                ::||||| :::|||:::
Db   182515 CCATCATCGTGACTTTTCCTCCAGACAATGAA-----ATGGGAGAAGAAAGTAGAGAC 182462

Qy   473 ValLeuThrAsnLeuLeuAsnAsnProSerLeuAlaAlaLeuPheMet 488
                ::: ||||||:::|||
Db   182461 ATCATGGACCAAGATTCGATGCTGATGCCTCCTCCTCCACCTCCACCACCT 182402

Qy   489 AspAspLysGluAsnThrLeuLysGlyArgGlnMetLeuThrAspGluGlnLysGly 508
                |||||||
Db   182341 ACAAATCAGTCCTATTTTTCAAACAACAACAACAATCCATCCTTTTTGAGTACAGTCGTT 182282

Qy   509   ArgThrArgValLysThrIleArgAlaLeuProArgLeuPhe 522
                        :::||||||||||||||:::
Db   182281 ---GGAAAGGTTAACAACACATTAGGTGGGAAGAGGCGAAGAAGATTACACAAGACT 182225

Qy   523 GlyAlaProThrAlaLys------------AlaGluMetIleAspAlaLysValPheGln 538
                |||||| |||||||||
Db   182224 ATGGAGTCTATAATCTTAAAAACAAGGTAAAAACTCTT--- 182186

Db   182185 ---CTAGAAACAACAAAGAACCTTCAATGCTCTGAACTGGTGAAAGTT---GTGTTCCAA 182132

```
CC   obesity, Crohn's disease, osteoporosis, inflammatory bowel disease,
CC   infertility, atherosclerosis, hypertension, scleroderma, haemophilia,
CC   diabetes, pancreatitis, autoimmune disease, asthma, arthritis,
CC   immunodeficiencies, HIV, viral, bacterial or parasitic infections, or
CC   graft-versus-host disease. The present sequence is a coding sequence of
CC   the invention
XX
SQ   Sequence 7396 BP; 1880 A; 2014 C; 2291 G; 1211 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:                            Length:          7396
Score:               0.000596         Matches:          170
Percent Similarity:  176.50           Conservative:      95
Best Local Similarity: 31.25%         Mismatches:       256
Query Match:         20.05%           Indels:           329
DB:                  5.36%            Gaps:              34
                     6
```

```
US-10-736-868-2 (1-643) x ABT08505 (1-7396)

Qy      12 GlyPheCysIleAlaProLeuSer------AlaGlnSerProSerThrSerAsp------ 27
              |||    ||     ||   :::    ||||||| ||:|||
Db    3999 GGCTTCTCAGCCAGTCCGACAGCCAAGTCCAGCCAAGGACTTCTCCCGCGCTGG 4058

Qy      28 AlaProGly------AlaLeuLeuSerSerLeuValGlyLysSerHisGlnLysLeuPro 45
              :::|||        :::  |||::   || ::|  :||:|
Db    4059 AGTCCCAGCTGCGAAGAGCACTCAGGAGCTG-CTGCAGGAGGAGAACCGGCAGAAGCTGAGC 4117

Qy      46 LeuAlaProSerMetGluAlaLeuGlu------------------------LeuAla 54
              ||:::         ::|||||
Db    4118 CTGAGCACCAAGCTCAAGCAGGTGAGGACGAGAGAGAATTCCTTCCGGGAGCAGTCGGAG 4177

Qy      55 ------------------------LeuMetGlyValGlnPheVal 61
                                       :::      |||
Db    4178 GAGGAGGAGGCCAAGCCAACCACCTGGAGAAGCAGATCGCCACCTCCAGGTGGCC 4237

Qy      62 AspAlaLeuIleLysGlyGlnMetGluMetAlaLysPheLysThrGlnLeu 81
              ||||||   :::   ||||||  ::: ||||||   |||||:
Db    4238 GAC------------ATGGAAAGAAGATGGAGGACAGTGTGGGGTGCCTGGAAACTGCTGAG 4288

Qy      82 GluValLeuGluLysValHisProAsp------------------------ 90
              |||||||   |||    ||||
Db    4289 GAGGTGAAGAGGAAGAGCTTCCAGAAGAGGACCTGAGGGCACGAGGAGAAG 4348

Qy      91 ------GlnPheAspLysTyrLysLeuLys------------ValAspAsp 102
                     ||         ||       |||           ||||||:
Db    4349 GTGGCCGCCTACGACAGCTGGAGGAGGACCAGGAGCGGCCTGCAGCAGGAGCTGGACGAC 4408

Qy     103 LeuAlaLysAspMetGlnGlnAlaLysGlnMetAlaLysLeuGlnLysProLysSerGly 122
              |||  :::|||    || |||  :||||||  ||||   ||||
Db    4409 CTGCTGGTGGTGACCTGGACCTGGACGCGCCTGCAACCTGAGAAGAAGCAGAGAG 4468

Qy     123 AsnAlaPheIleAspMetLeuAsnGlyAsnGlyIlePro IleGly------------ 137
              ||     :::       ||
Db    4469 AAG------TTTGACCAGCTTCCTGCCAGGAGGAGAACCATCTCTGCCAGTATGCAGAG 4522

Qy     138 ------------------------------------SerSerIle 140
                                                    |||
Db    4523 GAGGCGACCGCGGCTGAGGCGCCGAGAGGACGCCAAGGCTCGTCGGCC 4582

Qy     141 ArgGlyLeuGluAspAlaIleArgThrGlnIleArgAspMetGluAsnThrAspProSerGlu 160
              ||||||||  ::: |||::|||:::  |||
Db    4583 CGGGGACCCGGGGATGATGTCGAGACAGAAGGCGAGCGGGGCTCAAC------------ 4633

Qy     161 GlnIleIleAlaLysAlaValMetAspLysPheGlnIleLeuProGlyLeuValAla 180
              :::|||||||:::|||::||||:||
Db    4634 ------AAGCAGTTCCGACCACGGAG------ATGGAG 4657

Qy     181 AsnMetIleAlaGlyLysAsnProPhe------LysMetProGlnGlnMetArgLysAlaGln 199
              ||     :::       ||||||:  ::::
Db    4658 GACCCTTATGAGCTCCAAGGATGATGTCCACGAGTGTTCCACGAGGAGAAGTC-CAA 4716

Qy     200 AlaAlaProSerValPheGlnGlnAlaLeuAlaGlnArgLysLeuGlyLysLeuLysAsn 219
              ||||||  :::
Db    5758 GCTCAAGCGGGCAGCTGGAGGAGGAGGCCG---AAGAGGAGGCCAACGCGGCCTCCCG 5814
```

```
Db    4717 GCGGGGCCT------------------------------AGAGCACCAGGTGAGGAGGAGAT 4746

Qy     220 ------------AlaProValAlaGlyValArgGlyValGlyGluGlnArgMetMetAsn-Ar 236
                           |||||||
Db    4747 GAAGACCCAGCTGGAAGAGCTGGAGGAGCGAGCTGGAGGACGAGCTGGAGATGCCAAGCTCG 4806

Qy     236 gValAspGlnArgMetGln------------------GlnArgGluGlnGluGluGluGluAs 250
              |:::|||              ::::|||||||
Db    4807 GTTGGAGGTCAACCTGCAGGCCATGAAGCCCAGTTGAGCGGGGAACCTGCAGGGGCCGGGA 4866

Qy     250 pGluAspAspAspAspLeuGluAspValProArgArgArgSerSerAspGlyGl 270
              |||  :::::   || ::: ||  :::
Db    4867 CGAGCAGAGCGGAGGAGAAGAAGACGCTGTCAGACAGGTGCGGGGAGATGAGGCGAGA 4926

Qy     270 uProGlnSerGluAlaGlnHisGlnArgAspLeuAlaArgArgLeuLysSerSerPr 290
              |||                 ::: |||
Db    4927 GCTGGAGGACGAGGAGAAGCAGCGCTCGATGGCAGTGCGGCCGGAAGAAGCTGGAGAT 4986

Qy     290 oArgLeuLysGluLeu------------------------------------ 295
              :||||||||:
Db    4987 GGACCTGAAGGACCTGGAGGGCCACATCGACTCGGCCAACAAGAACCGGACAGAAGCCAT 5046

Qy     296 ------LeuGlnAsnAlaGluValGlnSerLeuLeuSerTyrGlnArgMetArgAspSe 313
                     ||| :::   ::: |||
Db    5047 CAAACAGCTGCGGAAGCTGCAGAGCTGCATGCGCGAGCTGGATGACAC 5106

Qy     313 rProLeuSerArgArgProLeuAlaMetAsnAspGluAspGluSerAlaPheArgAl 333
              : |||         |||::   ||||
Db    5107 CCGCGGCCTCTCGTGAGGAGATCCTGGCCAGGCCAAAGAGAACGGAGAAGAGCTGAAGAG 5166

Qy     333 aMetGluAla------------------------ArgAlaLys 339
              :||||||||     ::||||||||
Db    5167 CATGGAGGCCCAGATGATCCAGTTGCAGGAGGAGAACCTGCAGCCGCGTGCCAAGCG 5226

Qy     339 ------------------------------------------- 339
Db    5227 CCAGGCCCAGAGGAGCGAGGATGAGCTGGCTGACAGAGTCGGCAGACAGCAGCCAAAGG 5286

Qy     340 ------------LeuAspGlnLysSerGlnLeuValLeuGlyLeuHisGlyPheGlySe 356
                           |||::
Db    5287 AGCCCTGGCGTTAGAGGAGAAGCGGCGTCTGGAGGGCCATCCGCCCAGCTGGAGGAGGA 5346

Qy     356 rAspAspAspGluAspGlyAsnLeuIleLeuAspProSerGluAsnSerPheAr 376
              ::: |||              ::
Db    5347 GCTGAGGAGGAGGGCCAGGGCAACAGCGAGCTGATCAACGAGGCTGAAGAAGGCCAACCT 5406

Qy     376 gArg------------------------------AlaProLeuArgLeuSerSerGlyPheValGluValLysLeuLy 391
              :::                                :::|||||
Db    5407 GCAGATGACCAGATCAACGCCGAACCTGGAGCTGGCGCGCGGGCCACGCGCCAGAG 5461

Qy     391 sSerAsnAspGluLeuLeuLysSerAlaLeuAspArg---------------IleLysTyrArgVa 407
              ||               ||||||     :::
Db    5462 ---AACGAGAATGTCTCGGCAGCAGCTGGCAGCAGAACAAGGAGCTTAAGGTCAAGCT 5517

Qy     407 lAspAspValGlu------------------------LysTyrLeuAla-ProLysPro 418
              :|||||||||||                         ||||||
Db    5518 GCAGGAGATGGAGGGGCACTGTCAAGTCCATCACCGCCCTGCCCGGC 5577

Qy     419 ------------------MetGluPheAsnProLysPro---------------GlnProG 428
                                ::: ||||||                       :::
Db    5578 CAAGATTGCACAGCTGGAGGAGCTGGACAACGACCAAGGAGCGCCAGGCAGCCTG 5637

Qy     428 lyTyr-------------------------------------------------- 429
              |||
Db    5638 CAAACAGGTGCGTCGGACCGAGAAGAAGCTGAAGAGTGCTGCTCAGTGGATGACGA 5697

Qy     430 ------PheAlaProArgLysIleProThrArgProArgLysMetLeuProLeuLeuIleG 448
                     ||||||
Db    5698 GCGGGAGAAACGCCGAGCAGTACAGGCCGACAAGGCATCTACCCGCTGAAGCA 5757

Qy     448 lySerAspProLysValGlnGluGlnIleArgArgHisProSer------------------ 462
              :::                                ::
Db    5758 GCTCAAGCGGGCAGCTGGAGGAGGAGGCCG---AAGAGGAGGCCAACGCGGCCTCCCG 5814
```

```
DR   WPI; 2001-639362/73.
DR   P-PSDB; ABG01716.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity.
XX
PS   Claim 1; SEQ ID NO 1707; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC   sequences. (I) is useful as hybridisation probes, polymerase chain
CC   reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC   and in recombinant production of (II). The polynucleotides are also used
CC   in diagnostics as expressed sequence tags for identifying expressed
CC   genes. (I) is useful in gene therapy techniques to restore normal
CC   activity of (II) or to treat disease states involving (II). (II) is
CC   useful for generating antibodies against it, detecting or quantitating a
CC   polypeptide in tissue, as molecular weight markers and as a food
CC   supplement. (II) and its binding partners are useful for treating disorders
CC   of sites expressing (II). (I) and (II) are useful in medical imaging
CC   involving aberrant protein expression or biological activity. The
CC   polypeptide, forensics, gene mapping, identification of mutations
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic
CC   coding sequences of the invention. Note: The sequence data for this
CC   patent did not appear in the printed specification, but was obtained in
CC   electronic format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 6432 BP; 1682 A; 1699 C; 2071 G; 980 T; 0 U; 0 Other;
```

```
Alignment Scores:
Pred. No.:                         6432
Score:                0.00073      Length:              6432
Percent Similarity:   174.50       Matches:              167
Best Local Similarity: 35.04%      Conservative:         107
Query Match:          21.36%       Mismatches:           278
DB:                   5.30%        Indels:               232
                                   Gaps:                  32
```

```
US-10-736-868-2 (1-643) x AAS65903 (1-6432)

Qy      12 GlyPheCysIleAlaProLeuSer----AlaGlnSerProSerThrSerAsp------ 27
           |||||   :::           :::          ||||||||
Db    4106 GGCTTCTCAGCCAGTCCCGACAGCAGTCAGCAAGCTCACCAAGGACTTCCGGCGCTGG 4165

Qy      28 AlaProGly---------AlaLeuLeuSerSerLeuValGlyLysSerHisGlnLysLeuPro 45
           :::||                    |||:::||||       :::::: |||||||||
Db    4166 AGTCCCAGCTGCAGCAGGACACTCAGGAGCTG-CTGCAGGAGGAGGAACCGGGAGAAGCTGAGC 4224

Qy      46 LeuAlaProSerMetGluAlaLeuGlu------ 54
           ||| ::::          ::::::       |||
Db    4225 CTGAGCACCCAAGCTCAAGCAGGTGGAGGACGAGAGAATTCCTTCGGGAGCAGCTGGAG 4284

Qy      55 ----------LeuMetGlyValGlnPhe 60
                               ::: |||
Db    4285 GAGGAGGAGGAGGCCAAGCACCAACCTGAGAAGAGCAGATGCCACCCTCCATGCCCAGGTG 4344

Qy      61 ValAspAlaLeuIleLysGlyGlnMetGluLysAlaLysPheLysThrGln 80
           |||            :::|||||||||| ::: |||||||
Db    4345 GCCGAC---------ATGAAAAAGAAGATGGAGACAGTGTGGGGTGCCTGGAAACTGCT 4395

Qy      81 LeuGluValLeuGluLysValHisProAsp------ 90
           |||||||             |||:::|||
Db    4396 GAGGAGGTGAAGAAGCAGAAGTCCAGAAGGACCTGGAGGGCGTGCACCAGGAG 4455

Qy      91 ------GlnPheAspLysTyrLysLysLeuLys-----------ValAsp 101
                 :::|||             :::|||        |||:|||
Db    4456 AAGGTGGCCGCCCTACGACAAGCTGGAGAAGACCAAGACGCGGCTGCAGCAGGAGACTGGAC 4515

Qy     102 AspLeuAlaAlaAspAlaValMetGlnGlnAlaGluMetAlaLysAlaLysLeuGlnProLysSer 121
```

```
Qy     462 ------------------------------ 462
Db    5815 CCGGAAACTGCAGCGCGAGCTCGGAGGACGGCCACTGAGACGCCGATGCCATGAACCGCGA 5874

Qy     463 ------------------------------ThrG 464
Db    5875 AGTCAGCTCCTAAAGAACAAGCTCAGGCGGCGGGGGACCTGCCGTTGTGTGCCCGCCG 5934

Qy     464 luTrpLys------------------------IleAlaLysGluSerA 472
           |||||                               :::||||||
Db    5935 AATGGCCCGAAAGGCGCCGGGATGGCTCCGACGAAGAGGTAGATGGCAAAGCGGATGG 5994

Qy     472 rgValLeuThrAsnLeuLysAsnAsnProSerLeuAlaAlaLeuPheMetAspAspLysL 492
           ||||||           ||||||     |||||
Db    5995 GGCTGAGGCCAAACCTGCCGAATAAGCCTCTTCCTCCGAGCCTGAGATGGATGGACAGA 6054

Qy     492 euGluAsnThrLeuLysGlyArgMetLeuThrAspGluGlnLysGlyArgThrArgV 512
           ::: |||          :::  :::
Db    6055 CAGACACCACAGCCTCCCCTTCCAG----ACCCCGCAG------------------ 6089

Qy     512 aLysThrIleArgLeuProArgLeuPheGlyAla---------------ProThrA 527
                |||   :::|||                             |||
Db    6090 --------CACGCCTCTCCCCACCTTCTTGGACTGCTGTGAACATGCCTCCTCCTG 6138

Qy     527 laLysAlaGluMetIleAspAlaLysValPheGlnAspIleGluGluArgProIleProP 547
           ||||||           :::|||   :::|:::||
Db    6139 CCCTC----------------------------------------CGCCCCGTCCCCC 6156

Qy     547 roLeuPheGluProLysGlyArgHisThrArgLeuArgTrpThrArgGlyAlaAsnGluL 567
           |||                       :::
Db    6157 CATCCCGTTCCCT------------------------------------- 6170

Qy     567 ysGlyIleProGlyLeuGlySerArgPheIleLeuProSerLeuAspPro------T 584
           ||||||::: |||     |||||| ||| |||
Db    6171 ------CCAGGTGTTGTTGAGGGCATTTGGCTTCCTCTGCTGCATCCCCTTCCAGCTC 6222

Qy     584 hrMetProAlaLeuAsnThrAlaPheSerThrGlnGlyAlaArgAspGluTrpAspT 604
           |||                     |||||||  ::: |||||
Db    6223 CCTCCCCTGCTCAGAATCTGATACCAAAGAGACAGGGCCCGGGCCCAGGCAGGAGAGCGACC 6282

Qy     604 hrMetPheLysIleProAsn 610
           ||||||||
Db    6283 AGCAGGCTTCCTCAGCCCTCT 6302
```

```
RESULT 6
AAS65903
ID   AAS65903 standard; cDNA; 6432 BP.
XX
AC   AAS65903;
XX
DT   13-FEB-2002 (first entry)
XX
DE   DNA encoding novel human diagnostic protein #1707.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US008631.
XX
PR   31-MAR-2000; 2000US-00540217.
PR   23-AUG-2000; 2000US-00649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT, Liu C, Tang YT;
XX
```

```
Db  4516 GACCTGCTGGTGACCTGGAAGCGGCGTCCAGAGCGCGTGCAACCTGGAGAAGAGCAG 4575
             |||||                              |||::::            |||
Qy   122 GlyAsnAlaPheIleAspMetLeuAsnGlyAsnGlyIleProIleGly---------- 137

Db  4576 AAGAAG------TTTGACCAGCTCCTTGGCCGAGGAGAAGACCATCTCTGCCAAGTATGCA 4629
             ||||||
Qy   138 ------------------------------------------------SerSer 139

Db  4630 GAGGACGGCGACCGGGCTGAGGCGGAGGAGAAGAAGACCAAGGCTCTGTCGCTG 4689
Qy   140 IleArgGlyLeuGluAspGluAlaIleArgThrGlnAsnThrAspProSer 159

Db  4690 GCCCCGGCCCCTGGAGGCCCGGAGCTGGAGCGGCCTCAAC------ 4743
Qy   160 GluGlnIleAlaAlaLysAlaValAlaMetAspLysPheGlnIleLeuProGlyLeuVal 179

Db  4744 ---------------AAGCAGTTCCGCACGGAG 4764
             :::|||||||||||:::
Qy   180 AlaAsnMetIleAlaGlyLeuLysAsnProPhe------ 198

Db  4765 GAGGCAACCTTATGAGCTCCAAGGATGATGTGGGCAAGAGTGTCCACCAGCTGGAGAAGTC- 4823
Qy   199 GlnAlaAlaProSerValPheGlnGlnAlaLeuAlaGlnAlaMetLeuGlyLys------ 218

Db  4824 ----------------AGAGCAGGTGGAGGA 4853
             :::
Qy   219 Asn----------AlaProValAlaGlyGlyValGlyGluGluGlnAlaArgMetMetAsn 235

Db  4854 GATGAAGACGCAGCTGGAAGACTGAAGATGCCAAGCT 4913
Qy   236 -ArgValAspGlnArgMetGln----------GlnArgGluLeuGlnGlu 249

Db  4914 GCGGTTGGAGGTCAACCTGCGAGCCCCATGGCCCCATTCTGAGGCGGGACCTGCAGGGCCG 4973
Qy   249 GlnAlaAlaGlyProSerValPheGlnGlnAlaLeuArgArgSerSerAspGl 269

Db  4974 GGACGAGCAGCAGGAGAAGAAAGCAGCTGCTGGTCAGACAGGTGCGGGAGATGCGAGGC 5033
Qy   269 uAspGluAspAspAspLeuGlu---ValProArgArgArgSerSerAspGl 289

Db  5034 AGAGCTGGAGGACGAGAAGCAGCGCTCGATGGCAGTGGCCGCCGGAAGAAGTCTGGA 5093
Qy   289 yGluProGlnSerGluValAlaGluHisGlnHisArgArgAspAspLeuLysLysSerSe 289

Db  5094 GATGGAGACCTGGAAGGACCTGGAGCGGCGACCGGCCACCGGCCCACAACCGGGACCCAAGC 5153
Qy   289 rProGluLeuGluLeu------ 295

Db  5154 CATCAAACAGCTGCGGAAGCTGCGAGGACTGAACGACCTGCATGCGGCCTGCGATGA 5213
Qy   296 ------LeuGlnAsnAlaGluValGlnSerLeuLeuSerTyrGlnArgMetArgAs 312

Db  5214 CACCCGCGCCCTCGTTGAGGAGATCTGGCCCCAGGCCAAGAGACACGAGAAGAAGCTGAA 5273
Qy   312 pSerProLeuSerLysArgArgProLeuAlaMetAsnAspAspGluSerAlaPheAr 332

Db  5274 GAGATGATCCCAGTTGCAGGGAGGAGAAACTGCAGC 5318
Qy   332 gAlaMetGluAlaMetLysArgAlaLysLysAspAspLeuAspAspLeuVvalLeuGlyLeuHisGl 352

Db  5319 CGCGGAGCGTCCAAGCGCGTGCAGCGCCCCAGCAGGAGGCTGGCTCGACGAGATCGC 5378
Qy   352 yPheGlyGlyValGluSerAspAspGluGluGluAspGlyGluAsnLeuIleAspProSerGl 372

Db  5379 CAACAGCAGCGCCAAAGGAGGCGGCTGGCGTCTTAGAGAGAAGAAGGAGGCCCGCAT 5438
Qy   372 uAsnSerPheArgArgAlaProLeuLeuSer------ 383

Db  5439 CGCCCAGCTGGAGGAGCTGGAGGAGGAGGCAGGGCCAACACGAGCTGATCAACGACCG 5498
Qy   384 -SerGlyPheValGluLysLysLysSer----------AsnAspGl 395

                                                    ---------LeuAspArgIle
Qy   395 uLeuLysSerSerAla------ 403
```

```
Db  5499 GCTGAGAAGGCCAACCTGCAGATCGACCAGATCAACACCGACCTGAACCTGAGCGGCGAG 5558
Qy   404 ----------------LysTyrArgValAspAspProValGluLysTyrLeuAl 415

Db  5559 CCACGCCCAGAAGAACGAGAATGCTCGGCAGCCAGCTGGAACGGCCAGAACAAGGAGCTTAA 5618
Qy   415 aProLysProMetGluPheAsnProLysProGlyTyrPheAlaProArgLysIl 435

Db  5619 GGTCAAGCTGCGAGGAGGATGGAGGGCACTGTCAGTCCAAGTACAAGGCC------TCCAT 5672
Qy   435 ---------------------------------------------SerSer 455

Db  5673 CACCGCCCTGGAGCCAGATTGCACGATTGACACAGCTGGCACGAGACGACCAAGGA 5732
Qy   455 eProThrArgProArgLysMetLeuProLeuLeuIleGlySerAspProLysValGlnGl 455

Db  5733 GCGGCCAGCGCCTGCAAACAGCAGTGCCGCGG 5778
Qy   455 u--------GluIleArgArgHisProSerThrGluThrLysTrpLysIleAlaLy 469

Db  5779 ---AAGGATGTGCTGCTGCAGGTGGATGAGATGACGGAGGAACGCCGAGCAGTACAAGGA 5834
Qy   469 sGluSerArgValLeuThrAsnLeuLysLysAsnProSerLeuAlaAlaLeuPheMetAs 489

Db  5835 CCAGGCCAAGGCCATCTACCCGCTGCAGGCGCTCAAGCCGCAGCTGGACGGAGGCCGA 5894
Qy   489 p----AspLysLeuGluGluAsnThrLeuLys------GlyArgGlnMetLeuThrAspGl 505

Db  5895 AGAGGAGGCCCAGGCGCCAACGCCTCCCGCCCGGAAACTGCGAGCGC 5940
Qy   505 uGlnLysLysGlyArgThrArgValLysThrLeuArgArgValLeuThrArgAlaLeuProArgLeuProArgGlyLeuPheGlyAlaPr 525

Db  5941 ---------------GAGCTGGAGGACGCGCCACTGAGACGGCCATGAACCTGCGAGT 5987
Qy   525 oThrAlaLysAlaGluGluMetIleAspAspAlaLysLysValPheGlnAspIleGluGlnArgProIl 545

                                                         -----PhePheGluProLysSerGlyAr
Qy   545 ePheGluProLys-------- 555

Db  5988 CAGCTCCTAAAGAACAAGCTGCAGGCGGGGACCTGCCGTTGTCGTGCCCGCCGAAT 6047
Qy   555 ePheGluProLeu------ 575

Db  6048 GGCCCCGGAAAGGCGGCCGCCGGGATGGCTCCGACGAGAGCCCATCCGGAAGGTCATCATCCTC 6107
Qy   555 gHisThrArgLeuArgTrpThrGlyAlaAsnGlyLysGlyIleProGlyLeuGlySerAr 575

Db  6108 AAAGATCAAAGATAAAATTCATGAAGATAAGATAAGGAAAAACACAAAAACGCTGAAAAATTC 6167
Qy   575 -Se 580

Db  6168 CAAAGACCAGATGTCTCCTTCTCCAAATGATGTTGCAACTCTCTTCCAGCAAGGGCACAAAA 6227
Qy   580 rLeuAspProThrMetProAlaLeuAsnThrArgGlnMetArgGlyArgAs 600

Db  6228 C------TGG------ATGGAGAATGAATTGACGAATTTGACAGAAGTAGGCTT 6269
Qy   600 pGluTrpAspThrMetPheLysIleProAsnAsnTrpAsnProGlyEvalGlyIleValGlyAspGluGluGlyIleProGluGluPrp 620

Db  6270 CAGA 6273
             |||:
Qy   620 eLys 621
```

RESULT 7
ADN60282

KW   sugar transporter; tumourigenesis; antiangiogenic; cytostatic;
KW   cerebroprotective; vasotropic; antiinfertility; cardiant;
KW   antibody therapy; antisense therapy; RNA interference therapy;
KW   RNAi therapy; cancer; stroke; infertility; heart disease; human; gene;
KW   ss.
OS   Homo sapiens.
FH   Key   Location/Qualifiers
FT   CDS   1..5883
FT   /*tag= a
FT   /products "non-muscle myosin heavy chain 9"
XX
PN   WO2004039955-A2.
XX
PD   13-MAY-2004.
XX
PF   29-OCT-2003; 2003WO-US034281.
XX
PR   29-OCT-2002; 2002US-0421989P.
PR   17-OCT-2003; 2003US-0512251P.
XX
PA   (RIGE-) RIGEL PHARM INC.
XX
PI   Lorens JB, Atchison RE, Friera A, Holland S;
XX
DR   WPI; 2004-376181/35.
DR   P-PSDB; ADN60283.
XX
PT   Identifying a compound that modulates angiogenesis or tumourigenesis,
PT   useful in diagnosing and treating angiogenesis, cancer, stroke,
PT   infertility and heart disease, comprises contacting the compound with
PT   angiogenesis polypeptide.
XX
PS   Disclosure; Page 71-72; 105pp; English.
XX
CC   The present invention describes a method for identifying a compound that
CC   modulates angiogenesis or tumourigenesis. The method comprises: (a)
CC   contacting the compound with angiogenesis polypeptide, e.g. Axl, tubulin
CC   cofactor D, transglutaminase 2, cytosine deaminase, peptidase M41
CC   (paraplegin), CD13 aminopeptidase, PPK-1, zip kinase, Gas6, SRm160, non-
CC   muscle myosin heavy chain, calmodulin 2, novel symporter, novel
CC   semaphorin, novel zinc finger helicase (FLJ22611), plexin-A2,
CC   deoxycytidylate deaminase or novel sugar transporter; (b) determining the
CC   functional effector of the compound upon the angiogenesis polypeptide or
CC   the physical effect of the compound upon the target polypeptide or its
CC   fragment or inactive variant; and (c) determining the chemical or
CC   phenotypic effect of the compound upon a cell comprising the target
CC   polypeptide or its fragment or inactive variant, thus identifying a
CC   compound that modulates cell cycle arrest. Also described is a method of
CC   modulating angiogenesis in a subject. The angiogenesis or tumourigenesis
CC   modulating compound has antiangiogenic, cytostatic, cerebroprotective,
CC   vasotropic, antiinfertility and cardiant activities, and can be used in
CC   antibody, antisense and RNA interference (RNAi) therapies. The method is
CC   useful in identifying a compound that modulates angiogenesis. The methods
CC   and compounds or compositions are useful in diagnosing and treating
CC   angiogenesis, cancer, stroke, infertility and heart disease. The present
CC   sequence encodes a human non-muscle myosin heavy chain 9 protein, which
CC   is used in the exemplification of the present invention.
XX
SQ   Sequence 7274 BP; 1861 A; 1977 C; 2247 G; 1189 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:   0.00103   Length: 7274
Score: 173.50   Matches: 178
Percent Similarity: 30.71%   Conservative: 91
Best Local Similarity: 20.32%   Mismatches: 263
Query Match: 5.27%   Indels: 346
DB: 13   Gaps: 35

US-10-736-868-2 (1-643) x ADN60282 (1-7274)

Qy   12 GlyPheCysIleAlaProLeuSer------AlaGlnSerProSerThrAsp------ 27

Db   3860 GGCTTCTCAGCCAGTCGACCAGCAGTCAGTCAAGCTCACCAAGGACTTCTCCGGCCTGG 3919
           :::||| ||||||

Qy   28 AlaProGly-----AlaLeuLeuSerSerLeuValGlyLysSerHisGlnLysLeuPro 45

Db   3920 AGTCCCAGCTGCAGCACTCAGGAGCTG-CTGCAGGAGGAGAACCGGCAGAAGCTGAGC 3978
           :::||| ||||| ||||||

Qy   46 LeuAlaProSerMetGluAlaLeuGlu---------- 54

Db   3979 CTGAGCACCAAGTCAAGCAGGTGGAGGACGAGAAGAATTCCTTCGGGAGCAGCTGGAG 4038
           |||::

Qy   55 ------LeuMetGlyValGlnPhe 60

Db   4039 GAGGAGGAGGAGGCCAAGGACACAACCTGGAGAAGCAGATCGCCACCTCCATGCCCAGGTG 4098
           :::

Qy   61 ValAspAlaLeuIleLysGlyGlnMetGluMetAlaLysGlyAlaPheLysThrGln 80

Db   4099 GCCGAC-----ATGAAAAGAAGATGGAGCAGTGGGGTGCCTGGAAACTGCT 4149
           || ::::||||| ::: ||

Qy   81 LeuGluValLeuGluLysValHisProAsp----------- 90

Db   4150 GAGGAGGTGAAGAGGAGTCCAGAAGACACCTGGAGGGCCTGAGCCAGCGGCACGAGGAG 4209
           ||

Qy   91 ------GlnPheAspLysTyrLysLysLeuLys 101

Db   4210 AAGGTGGCCGCCTACGACAAGCTGGAGAAGACCAAGAACCCGGCTGCAGCAGGAGCTGGAC 4269
           :::|||||| ||||||

Qy   102 AspLeuAlaAlaAspAlaValMetGlnHisAlaGluMetAlaLeuLysLeuGlnLysLysSer 121

Db   4270 GACCTGCTGGTGGACCTGGACCACCAGCCCCAGGCGCCGTGCAACCTGGAGAAGAGCAG 4329
           |||||| :::

Qy   122 GlyAsnAlaPheIleAspMetLeuAsnGlyAsnGlyIleProIleGly----------- 137

Db   4330 AAGAAG------TTTGACCAGTCCTCTGGCCGGAGGAAGACCATCTCTGCCAAGTATGCA 4383
           ||||||

Qy   138 ------SerSer 139

Db   4384 GAGGAGGACCGCGGTCGAGGCGAGGCGGAGGCCCGAGAGAAGGAGAAGCCAAGGCTCTGTCGCTG 4443
           |||

Qy   140 IleArgGlyLeuGluAspArgThrGlnArgThrGlnArgAspMetGluAsnThrAspProSer 159

Db   4444 GCCCGGGCCCCTGGAGAGCCATGGAGCAGTGGAGCGGGAGCTGGAGCGGGCCTCAAC- 4497
           |||||||:::|| :::::::   :::

Qy   160 GluGlnIleAlaLysAlaValMetAspLysPheGlnThrGlnIleLeuLeuProGlyLeuVal 179

Db   4498 ------AAGCAGTTCCCCACCGAG------ATG 4518
                                              :::

Qy   180 AlaAsnMetIleAlaGlyLysAsnProPhe--LysMetProGlnMetArgLysAla 198

Db   4519 GAGGACCTTATGAGCTCCAAGGATGATGTGGGCCAAGAGTGTCCACGAGTGGAGAAGTC- 4577
           :::::::::::::::   :::   :::::::

Qy   199 GlnAlaAlaProSerValPheGlnGlnAlaLeuAlaLeuAlaGlnAlaLeuArgAlaMetLeuGlyLys 218

Db   4578 ------AGAGCAGCAGGGTGGAGGA 4607
                 |||

Qy   219 Asn------AlaProValAlaGlyValGlyArgGlyGluGlyGlnGlnArgMetMetAsn 235

Db   4608 GATGAAGACGCAGCTGGAAGACGAGCTGCAGGGACCACCGAAGATGCCAAGCT 4667
           :::

Qy   236 -ArgValAspGlnArgMetGln------GlnArgGlyLeuGlnGluGly1 249

Db   4668 GCCGTTGAGGTCAACCTGCAGGCCATGAAGGCCCAGTTCGAGCGGGACCTGCAGGGCCG 4727
           ::::::   :::|||

Qy   249 uAspGluAspAspAspProLeuGluAspGluAspValProArgArgSerSerAspG1 269

Db   4728 GGACGAGCAGGAGCGAGGAGAAGAAGACAGCTGGTCAGACAGTGCGGGAGATGGAGC 4787
           |||||| :::   :::

Qy   269 yGluProGlnSerGlyHisAlaGluHisGlnArgArgAspLeuArgArgLeuLysSerE 289

Db   4788 AGAGCTGGAGGACGAGAGGAAGCAGCAGCCGCCTCGATGGCGAGTGGCGCCCCCGGAGAAGCTGGA 4847
           :::|||

Qy   289 rProArgLeuLysGluLeu------ 295
           |||||:::||

```
Db   4848 GATGGACCTGAAGGACGCGCACATCGACTCGGCCCAACATGAACCGGGACGAAAGC 4907
Qy    296 ------------LeuLeuLysGlnAsnAlaGluValValGlnSerLeuLeuSerTyrGlnArgMetArgAs 312
Db   4908 CATCAAACAGCTGCGGAAGCTGCAGGCCCAGATGAAGGACTGCATGCGGCGAGCTGGATGA 4967
Qy    312 pSerProLeuSerLysArgArgProLeuAlaMetAsnAspGluAspGluSerAlaPheAr 332
Db   4968 CACCCGGCGCCTCTCGTGAGGAGATCCTGGCCCAGGCAAAGAGAGAACGAGAAGAAGCTGAA 5027
Qy    332 gAlaMetGluAlaArgAlaLysAspLeuAlaAspLysLysSerGlnLysLysValLeuGlyLeuHisGl 352
Db   5028 GAGCATGGAGGCC                                          GAGATGATCCAGTTGCAGGAGGAACTGGCAGC 5072
Qy    352 yPheGlyGluSerAspAspAspGluGluIleLeuProGlyLeuGlyLeuGlySerArgPheLeuAspProSerGl 372
Db   5073 CCGGGAGCGTGCCAAGCGCCAGCCAGGGCGGGATCGAGCTCGGCTGCGACGAGATCGC 5132
Qy    372 uAsnSerPheArgArgAlaProLeuSer                          ThrMetProAlaLeuAsnThrAlaPheSerThrGlnGly 383
Db   5133 CAACAGCAGCGGCAAAGGAGCCCTGCCGTTAGAGGAGAAGGCCGTCTGGAGGCCGCCAT 5192
Qy    383                                          ArgAlaArgAsp 600
Db   5193 CGCCCAGCTGGAGGAGGAGGAGGAGCAGGGCCAACACGAGCTGATCAACGACCG 5252
Qy    384 -SerGlyPheValGluLysLysSer                          AsnAspGl 395
Db   5253 GCTGAAGAAGGCCAACCTGCAGATCCACCAGATCAACACCCGACCTGGAGCCGCGCAG 5312
Qy    395 uLeuLysSerArgAla                          IleLyLeuAspArg 399
Qy    400 ------------------------------LeuAspArg------------IleLeLy 404
Db   5313 CCACCGCCCAGAAGAACGAGATGCTCGGCAGCAGCTGGAACCGCTGGAACAAGAAGAGCTTAA 5372
Qy    404 sTyrArgValAspAspValGlu                          LysTyrLeuAla-ProLysPro- 418
Db   5373 GGTCAAGCTGCAGGAGATGGAGGCCACTGTCAAGTCACAGGGCTACACAGGGCTCCATCACCGC 5432
Qy    419 ------------------------------MetGluPheAsnProLysPro 425
Db   5433 CCTCGGAGGGCCAAGATTGCACAGTCGAGGAGCAGCTGGACAACGAGGAGCAGGAGCAGGCCA 5492
Qy    426 --GlnProGlyTyr 429
Db   5493 GGCAGCTGCAAACAGGTGCGTCGAGCAGGAGAAGAGAAGGATGTGTCGTGCTGCAGGT 5552
Qy    430 ------------PheAlaProArgLysIleProThrArgProArgLysMetLeuProL 445
Db   5553 GGATGACGCAGCGGAGGAACCGCGAGCAGTACAAGGACCAGGCCGACAAGGCCATCTACCCG 5612
Qy    445 euLeuIleGlySerAspProLysProValGlnGluGluIleGluGlnArgHisProSer- 462
Db   5613 CCTGAAGCAGCTCAAGCGGCAGCTGGAGGAGGCCG--AAGAGGAGGCCCAGCGCGGGCCAA 5669
Qy    462 ------------------------------AsnSerAsnProSerLeuAlaAlaLeuPheMetA 469
Db   5670 CGCCTCCCGCGCCGAAACTGCAGCGCGGGAAACTGCAGCGCGCTGAGCGGCTGGAGGACGGCGGCCAT 5729
Qy    462                                          IleAlaL 469
Db   5730 GAACCGCGAAGTCAGCTCAGCTCCCTAAAGAACAACAGCTCAGGCGGCGGGGACCTGCGCGTTTGTCGT 5789
Qy    463 ----------------------------ThrGluTrpLys 469
Db   5790 GCCCCGGCGGAATGGGCCCCGGAAAGGCCGCCGGGATGGCTCCGACGAAGAGTAGATGGCAA 5849
Qy    469 ysGluSerArgValLeuThrLeuThrAsnLeuLysAsnProSerLeuAlaAlaLeuPheMetA 489
Db   5850 AGCCAGGATGGGGCTGGGCCAAACTGCCGAATAAGCCTCTTCTCCTGCAGCCTGAGATGG 5909
Qy    489 spAspLysLeuGluAsnThrLeuLysGlyArgGlnMetLeuThrPheAspGluGlnLysGlyLysA 509
Db   5910 ATGGACAGACAGACACCAGCCTCCCCTTCCCAG------ACCCCGACG 5953
```

```
Qy    509 rgThrArgValValThrIleArgArgAlaLeuProArgLeuPheGlyAla------------ 524
Db   5954 ------------------------CACGCTGCTCCCCACCTTCTTGGACTGCTGTGAACATGC 5993
Qy    525 --ProThrAlaLysAlaGluMetIleAspAlaLysValPheGlnAsnAspIleGluUArgP 544
Db   5994 CTCCTCCTGCCCTC------------------------------CGCC 6011
Qy    544 rolleProProLeuPheGluProLysGlyArgHisThrArgLeuArgTrpThrGlyA 564
Db   6012 CCGTCCCCCATCCCGTTCCCT 6034
Qy    564 laAsnGluLysGluIleProGlyLeuGlySerArgPheIleProSerLeuAspPro- 583
Db   6035 ------------CCAGTGTTGTTGTGAGGGCATTTGGCTTCCTCTGCATCCCC 6077
Qy    584 ------------ThrMetProAlaLeuAsnThrAlaPheSerThrGlnGly-ArgAlaArgAsp 600
Db   6078 TTCCAGCTCCCTCCCCTGCTCAGAATCTGATACCAAAGAACAGGGCCCGGGCCCAGGCA 6137
Qy    601 GluTrpAspThrMetPheLysIleProAsn------------ 610
Db   6138 GAGAGCGACCAGCAGGCTCCTCAGCCCTTCTTGCCAAAAGCACAGATGTTGAGGCGA 6197
Qy    611 ---AsnTrpAsnProGlyAspGluValGlyPheLysMetAsn 623
Db   6198 GCAGGGCCAGGCCCCCGGGGAGGGGGCCCAGAGTTTTCTATGAAT 6239
```

RESULT 8
```
AAK51987
ID   AAK51987 standard; cDNA; 7516 BP.
XX
AC   AAK51987;
XX
DT   06-NOV-2001  (first entry)
XX
DE   Human polynucleotide SEQ ID NO 532.
XX
KW   Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW   vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW   tissue growth factor; immunomodulatory; cancer; leukaemia;
KW   nervous system disorder; arthritis; inflammation; ss.
XX
OS   Homo sapiens.
XX
PN   WO200157190-A2.
XX
PD   09-AUG-2001.
XX
PF   05-FEB-2001; 2001WO-US004098.
XX
PR   03-FEB-2000; 2000US-00496914.
PR   27-APR-2000; 2000US-00560875.
PR   20-JUN-2000; 2000US-00598075.
PR   19-JUL-2000; 2000US-00620325.
PR   01-SEP-2000; 2000US-00654936.
PR   15-SEP-2000; 2000US-00663561.
PR   20-OCT-2000; 2000US-00693325.
PR   30-NOV-2000; 2000US-00728422.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT,   Liu C,    Drmanac RT,  Asundi V,   Zhou P,   Xu C,    Cao Y;
PI   Ma Y,  Zhao QA,   Wang D,  Wang  J,   Zhang J,    Ren F,    Chen R,   Wang ZW;
PI   Xue AJ,  Yang Y,  Wejhrman T,    Goodrich R;
XX
DR   WPI; 2001-476283/51.
DR   P-PSDB; AAM78854.
XX
PT   Nucleic acids encoding polypeptides with cytokine-like activities, useful
PT   in diagnosis and gene therapy.
XX
```

```
PS
XX   Claim 1; Page 1918-1926; 6221pp; English.
CC   The invention relates to polynucleotides (AAM78323-AAM80302) that exhibit activity elating to
CC   encoded polypeptides (AAK51456-AAK53435) and the
CC   cytokine, cell proliferation or cell differentiation or which may induce
CC   production of other cytokines in other cell populations. The
CC   polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC   peptide therapy. The polypeptides have various cytokine-like activities,
CC   e.g. stem cell growth factor activity, haematopoiesis regulating
CC   activity, tissue growth factor activity, immunomodulatory activity and
CC   activin/inhibin activity and may be useful in the diagnosis and/or
CC   treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC   inflammation. Note: Records for SEQ ID NO 2110 (AAK52581), 2111
CC   (AAK52582) and 3666 (AAM80020) are omitted as the relevant pages from the
CC   sequence listing were missing at the time of publication
XX
SQ   Sequence 7516 BP; 1909 A; 2056 C; 2326 G; 1225 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:              0.00107        Length:          7516
Score:                 173.50         Matches:         178
Percent Similarity:    30.71%         Conservative:    91
Best Local Similarity: 20.32%         Mismatches:      263
Query Match:           5.27%          Indels:          346
DB:                    4              Gaps:            35

US-10-736-868-2 (1-643) x AAK51987 (1-7516)
```

```
Qy   12  GlyPheCysIleAlaProLeuSer------AlaGlnSerProSerThrSerAsp------       27
Db 4094  GGCTTCTCAGCCACGAGTCCAGCACGCAAGTCCAGCAAGTCCTCACCCAAGGACTTCTCCGCGCCTGG  4153

Qy   28  AlaProGly------AlaLeuLeuSerLeuValGlyLysGlySerHisGlnLysLeuPro         45
Db 4154  AGTCCCAGCTGCGAGACACTCAGGAGCTG-CTGCAGGAGGAGAACCGGCAGAAGCTGAGC        4212

Qy   46  LeuAlaProSerMetGluAlaLeuGlu-------                                   54
Db 4213  CTGAGCACCACCAAGCTCAAGCAGGTGGAGGAGGAGAGAATTCCTTCCGGCAGCAGCTGGAG      4272

Qy   55  -------...----LeuMetGlyValGlnPhe                                     60
Db 4273  GAGGAGGAGGAGGCCCAAGCACAACCTGGAGAAGCAGCAGATCGCCACCCTCCATGCCCAGGTG    4332

Qy   61  ValAspAlaLeuIleLysLysGlyGlnMetGluMetAlaLysGlyAlaAlaPheLysAlaPheThrGln 80
Db 4333  GCCGAC------...------ATGAAAAAGAAGATGGAGGACAGTGTGGGTGCCTGGAAACTGCT    4383

Qy   81  LeuGluValLeuGluLysValHisProAsp-------                                90
Db 4384  GAGGAGGTGAAGAGAGAAGCTCCAGAGGAACCTGGAGGGCCTGAGCCGGCCACGAGGAG        4443

Qy   91  --------...----GlnPheAspLysTyrLysLysLeuLys------ValAsp              101
Db 4444  AAGGTGGCGCCGCTTACGACAAGCTGGAGAAGACCGGCTGCAGCAGGAGCTGAC            4503

Qy  102  AspLeuAlaAlaAspAlaValMetGlnAlaMetAlaLysLeuMetAlaLysGluMetGlnAlaLeuGlnProLysSer 121
Db 4504  GACCTGGCTGGTGACCTCGCCGCTGACCTCGCCGCTGACCTGGAGAAGAAGCAGCAG         4563

Qy  122  GlyAsnAlaPheIleAspMetLeuAspMetLeuAsnGlyAsnGlyIleProIleGlyGly--------  137
Db 4564  AAGAAG------...----TTTGACCAGCTCCTGGGCGGAGGAGAAGACCATCTCTGCCAAGTATGCA  4617

Qy  138  -------...------SerSer                                              139
Db 4618  GAGGAGGCGCCGACCGGGCTGGCGGGGCTCGAGCGGCGGAGGCCCGAGAGAGAGGACCAAGGCTCTGTCGCTG 4677

Qy  140  IleArgGlyLeuGluAspAlaIleArgThrGlnAspMetGluAsnThrArgAspProSer        159
Db 4678  GCCCGGCGCCCTGAGGAGGCCATGGAGGCCAGCAGAGGCGCAGAGGCGAGCCGCTCAAC-----    4731

Qy  160  GluGlnIleIleAlaLysAlaValAlaMetAspLysPheGlnThrGlnIleLeuProGlyLeuVal  179
Db 4732  -------...-----AAGCAGTTCCGCACGGAG                                     
                      ----ATG  4752

Qy  180  AlaAsnMetIleAlaGlyLysAsnProPhe---LysMetProGlnMetArgLysAla          198
Db 4753  GAGGACCTTATGAGCTCCAAGATGATGATGTGGCAGAGAGTGTCCACGAGCTGGAGAGTC-      4811

Qy  199  GlnAlaAlaProSerValPheGlnAlaLeuAlaGlnArgAlaMetLeuGlyLys             218
Db 4812  AGAGCAGCAGGTGGAGGA                                                  4841
         CAAGCGGCGCCCT------

Qy  219  Asn-------AlaProValAlaGlyAlaGlyGluGluGlnArgMetMetAsn              235
Db 4842  GATGAAGACAGCAGCTGGAGGACGAGCTGGAGGACGAGCTGCAGGCCACCGAGAATGCCAAGT    4901

Qy  236  -ArgValAspGlnArgMetGln-----------GlnArgGluLeuGlnGluGlu1           249
Db 4902  GCCGGTTGGAGGTCAACCTGCAGGCCATGAAGCCCAGTTCGAGCCGACCTCAGGGGCG        4961

Qy  249  uAspGluAspAspAspAspLeuGluGluAspGl...ArgSerSerAspG1               269
Db 4962  GGACCAGCAGCAGCGAGGAGAAGAGAAGCAGCTCGTCAGACAGGTGCGGGAGATCGAGGC      5021

Qy  269  yGluProGlnSerGluAlaGluHisGlnArgArgAspLeuAlaArgArgLeuLysSerSe     289
Db 5022  AGAGCTGGAGGACGAGGAGAAGCAGCGCGTCGATGGCAGTGCCCGGAAGAGAGCTGA        5081

Qy  289  rProArgLeuLysGluLeu-                                              295
Db 5082  GATGGACCTGGAAGGACCTGGAGGCGCACATGACCTCGGCCCAACAAGAACCGGACGAAGC     5141

Qy  296  ------LeuGlnAsnAlaGluValGlnSerLeuLeuSerTyrGlnArgMetArgAs         312
Db 5142  CATCAAACAGCTGCGGAAGCTGCAGGCCCCAGATGAAGGACTGCATGCGCCAGCTGGATGA     5201

Qy  312  pSerProLeuSerLysArgArgProLeuAlaMetAsnAspGluSerAlaPheAr          332
Db 5202  CACCCGCGCTCTCGTGAGGAGATCCTGGCCCAGGCCAAAGAGACGAGAAGAGCTGAA        5261

Qy  332  gAlaMetGluAlaArgAlaLysLeuAspLeuLysSerGlnLysLeuValLeuGlyLeuHisG1  352
Db 5262  GAGCATGGAGGCC------...----GAGATGATCCAGTTGCAGGAGGAACTGGCAGC       5306

Qy  352  yPheGlyGluSerArgAspAspGluAspGluGlnAspAsnLeuIleAspProSerG1       372
Db 5307  CGGCGAGCTGCCAAGCGCCAGGCGCCAGGAGGAGGGCGATGAGCTGGCTGCGACGATCGC     5366

Qy  372  uAsnSerPheArgArgAlaProLeuArgLeuSer-                             383
Db 5367  CAACAGCAGCGGCCAAAGGAGCCCTGGGGTTAGAGGAGGAAGCGGCGTCTGGAGGCCCGCAT   5426

Qy  384  -SerGlyPheValGlyLeuLysLysSer-------AsnAspGl                     395
Db 5427  CGCCCAGCTGGACGGAGCTGGAGGAGGAGGAGCAACACGGAGCTGATCAACGACACCG      5486

Qy  395  uLeuLysSerAla-------                                            399
Db 5487  GCTGAAGAGAGGCCCAACCTGCAGATCGACCAGATCAACACCGCACCTGAACCTGGAGCAG    5546

Qy  400  -----...----LeuAspArg--------IleLy                              404
Db 5547  CCACGCCCAGAGAACGAGAATGCTCGGCGCAGCTGCAGAACGAAGAGCTTAA           5606

Qy  404  sTyrArgValAspAspValGlu-------LysTyrLeuAla-ProLysPro-            418
Db 5607  GGTCAAGCTGCAGGAGATGGAGGGCACTGTCAAGTCCAATCCATCACCGC             5666

Qy  419  -------...------MetGluPheAsnProLysPro-------                    425
Db 5667  CCTCGAGGCCAAGATTGCACAGCTGGAGGAGCAGCTGGACAACCAGAACCGAGACGCGCCA    5726

Qy  426  --GlnProGlyTyr-------                                           429
```

```
XX
PD   16-MAR-2004.
XX
PF   30-SEP-1999;   99US-00409800.
XX
PR   30-SEP-1999;   99US-00409800.
XX
PA   (WISC ) WISCONSIN ALUMNI RES FOUND.
XX
PI   Blatner FR, Burland V, Rose DJ, Mayhew GF, Perna N, Perry RD;
PI   Straley SC, Fetherston JD, Lindler LE, Plano GV;
XX
DR   WPI; 2004-236724/22.
XX
PT   New recombinant DNA construct comprising an open reading frame placed
PT   under the control of a non-native promoter, useful for preparing a
PT   composition for diagnosing, treating or preventing infection caused by
PT   Yersinia pestis.
XX
PS   Example; SEQ ID NO 2; 90pp; English.
XX
CC   The invention relates to a recombinant DNA construct comprising an open
CC   reading frame placed under the control of a non-native promoter. The open
CC   reading frame comprises bp 2389-2826 of the Yersinia pestis plasmid
CC   pPCP1, representing the pesticin immunity protein, a bacteriocin. Also
CC   disclosed are the sequences of the Yersinia plasmids pCD1 (responsible
CC   for the complex virulence response termed LCR, low calcium response)and
CC   pMT1 (encoding capsular proteins and a murine toxin). The recombinant DNA
CC   construct is useful for preparing a composition for diagnosing, treating
CC   or preventing infection caused by Yersinia pestis (e.g. bubonic plague).
CC   The present sequence is the pMT1 plasmid.
XX
SQ   Sequence 100990 BP; 26256 A; 24506 C; 26153 G; 24075 T; 0 U; 0 Other;
```

```
Alignment Scores:
Pred. No.:                               Length:        100990
Score:                 0.0423  171.00    Matches:          142
Percent Similarity:    33.99%            Conservative:     100
Best Local Similarity: 19.94%            Mismatches:       260
Query Match:            5.20%            Indels:           210
DB:                    12                Gaps:              33

US-10-736-868-2 (1-643) x ADJ94407 (1-100990)
```

```
Db   5727 GGCAGCCTGCAAACAGTGCCTCGGACCGAGAGAGCTGAAGGATGTGCTGCTGCAGGT 5786
Qy    430 ----------PheAlaProArgLysIleProThrArgProArgLysMetLeuProL 445
               ::: |||||||
Db   5787 GGATGACGAGCGGAGGAACGCGAGCAGTACAAGGACCAGCCGACAAGGCATCTACCCG 5846
Qy    445 euLeuIleGlySerAspProLysValGlnGluGluIleuProSer- 462
               ::: ||||
Db   5847 CCTGAAGCAGCTCAAGCGGCAGCTGAGGAGGCCG---AAGAGGAGCCCAGCGGGCCAA 5903
Qy    462 --- 462
Db   5904 CGCCTCCCGCCGGAAACTGCGACGGCGAGCTGGAGGACGCGCCACTGAGACGGCCGATGCCAT 5963
Qy    462 --- 462
Db   5964 GAACCCGGAAGTCAGCTCCTAAAGAACAAGCTCAGGCGCGGGGACCTGCCGTTTGTCGT 6023
Qy    463 ----ThrGluTrpLys------IleAlaL 469
               ||||  ::: |||
Db   6024 GCCCCGCGAATGGCCCGGGAAAGGCGCCGGGATGGCGTCCGACGAGAGGTAGAATGGCAA 6083
Qy    469 ysGluSerArgValLeuThrAsnLeuLysAsnAsnProSerLeuAlaLeuAlaLeuPheMetA 489
               ::: |||  ||| |||
Db   6084 AGCGGATGGGGCTGAGGCCAAACCTGCCGAATAAGCCTCTTCTCCTGCAGCCTGAGATGG 6143
Qy    489 spAspLysGluAsnThrLeuLysLysGlyAlaGlnMetLeuThrHisAspGluGlnLysGlyA 509
               ::: ||||||| |||
Db   6144 ATGGACAGACAGACACCACAGCCTCCCTTCCAG------ACCCCGCAG--- 6187
Qy    509 rgThrArgValLysThrIleArgAlaLeuProArgAspPheGlyAla-- 524
               ||| |||||
Db   6188 ---CACGGCTCTCCCCACCTTCTTGGGACTGCTGTGAACATGC 6227
Qy    525 --ProThrAlaLysAlaGluMetIleAspAlaLysAlaValPheGlnAspIleGluGluArgP 544
               ||| |||
Db   6228 CTCCTCCTGCCCTC------------CGCC 6245
Qy    544 roIleProProLeuPhePheGluProLysGlyArgHisThrArgLeuArgTrpGlyThrGlyA 564
               ||:::|||||  |||
Db   6246 CCGTCCCCCATCCCGTTCCCT--- 6268
Qy    564 laAsnGluLysGlyLeuIleProGlyLeuGlySerArgPheIleLeuProSerLeuAspPro- 583
               ||||  ::: |||
Db   6269 ---CCAGTGTTGTTGAGGGGCATTTGGCTTCCTTCCTGCTCCGCATCCCC 6311
Qy    584 ------ThrMetProAlaLeuAsnThrAlaPheSerThrGlnGly-ArgAlaArgAsp 600
               ||| |||||||:::
Db   6312 TTCCAGCTCCCCTCCCCTGCTCAGAATCTGATCACCAAAGAGACAGGGGCCCCCAGGCA 6371
Qy    601 GluTrpTrpAspThrMetPheLysIleProAsn------ 610
               ||||  ::: |||
Db   6372 GAGAGCGACCAGCAGGCTCCTCAGCCCTCTCTGCCAAAAGCACAGATGTTGAGGCGA 6431
Qy    611 ---AsnTrpAsnProGlyAspGluValGlyPheLysMetAsn 623
               ||||||:::
Db   6432 GCAGGGCAGGGCCCCCGGGGGAGGGGAGCCCAGAGTTTTCTATGAAT 6473
```

```
Qy     17 ProLeuSerAlaGlnSerProSer---ThrSerAspAlaAspGlyAlaAlaLeuLeuSerSer 35
               |||||||||  ||||
Db  85225 CCAGTGCTGGTGAAACATCCCATCTATACCGGACGACGCATCACCACCAACCCTCATGAACGCA 85284
Qy     36 LeuValGly------ 38
               |||
Db  85285 GTACGTGGTTTTCTCGACCATGAAGTAGCCCATATCCTCTTTACTGATCCGAAAGTCGCG 85344
Qy     39 ------LysSerHisGlnLysProLeuAlaProSerMetGluAlaLeuGluLeu 55
               ::: ||||||
Db  85345 ATGAAGATGCGCGAGAGAGGAGGAAGGCTCCGTCTACCGGGCTTTGGAACGCGCTGGAA--- 85401
Qy     56 MetGlyValGlnPheValAspAlaLeuIleLeuLysLysGlyGlnMetGluAlaLysGly 75
               ||||||  :::
Db  85402 ------GACGTTTTATTGAACGACCAGAATGGGACAGGTATTCAACGGA 85443
Qy     76 AlaPheLysThrGlnLeuGluValLeuGluLysValHisProAspPheAspLysTyr 95
               |||  ||||
Db  85444 ACCCGTCGTAATTTGCTAGCCACAGAACCTGGTG------ATCGACAAATAC--- 85491
Qy     96 LysLysLeuLysValAspAspLeuAlaAlaAspAlaValMetGlnAlaGlnMet--- 114
               ||||||| :::
Db  85492 TTCAAGGGCAAAGTGTCAGAGGCGTTTCAATCTGCCACGGCAACCAGCCGTGAATTGTTC 85551
Qy    115 ------AlaLysLeuGlnProLysSerGlySerAla--PheIleAspAsp 127
               :::  :::
Db  85552 CTGAAATTCTTCCTTGCCCGGTGCGTCCGGACGGATGGCCAAAGCCCCTTTCATCGAC 85611
Qy    128 MetLeuAsnGlyAsnGlyIleProIleGlySerSerIle------ArgGly 142
```

```
RESULT 9
ADJ94407
ID   ADJ94407 standard; DNA; 100990 BP.
XX
AC   ADJ94407;
XX
DT   03-JUN-2004 (first entry)
XX
DE   Yersinia pestis plasmid pMT1.
XX
KW   Bubonic plague; ds; plasmid; cyclic; circular; pMT1; capsular protein.
XX
OS   Yersinia pestis.
XX
PN   US6706522-B1.
```

Left column

```
Db  85612 TTTATGGAAGAACACTGGCACCTCATCGATAAGGCTGTAGCCTTGCCTTGCTCAAAGAGCATGGT 85571
                :::                                    :::                     |||
Qy    143 LeuGluAspAlaIleArgThrGlnArgAspMetGluAsnThrAspProSerGluGlnIle 162
Db  85672 ATCGACGTGCCGGTT------CGCAATATGTCGAACACAGAGGACTGCGGTAAAGGTC 85722
                ::::::|||||::|:          ||||:::::::
Qy    163 AlaLysAlaValMetAspLysPheGlnIleLeuProGlyLeuValAlaAsnMet 182
Db  85723 GCAGCTGCTATCGCCCAAATCATGCAGGAT------ 85752

Qy    183 IleAlaGlyLysAsnProPheLysMetProGlnMetArgArgLysAlaGlnAlaAlaPro 202
                :::            |||||||:::            :::
Db  85753 ATGAAAAGACAAGCCAGAGGGCAAATTACCT--GAGCTTAAATCATCTGCCAAAAAGCCG 85809

Qy    203 SerSerValPheGlnGlnAlaLeuAlaLeuGlnAlaMetLeuGlyLysAsnAlaProVal 222
                |||                                    :::          |||
Db  85810 TCTAAGGACAAAGACAGTCC--------------------GAGAAACACCAGAG 85845

Qy    223 AlaGlyGlyValGlyGluGlnLeuGlnArgMetMetAsnArgValAspGlnArgMetGln 242
                :::|||          |||||||:::
Db  85846 TCAGGCGGACGATCCGTCTCATAGTGAATCTGCACCAAGCCGAACCAAAGGCGAAGACGAC 85905

Qy    243 GlnArgGluLeuGlnGluAsp---GluAspAspAspAspAspLeuGluLeuAspAspVal 261
                :::||||||||::|:                    ||||||:::
Db  85906 GACAAGGAAGAGCAAGAAGACATGATGAGTCCAGAGGAAGGATTCTGGGATTCTGATTTG 85965

Qy    262 ProArgArgArgSerArgSerAspGlyGlnSerGluGlnAlaGlyHisGlnArgArgAsp 281
                                    :::       |||
Db  85966 CCT------GATTCATTAGATAAGGACTTACCTATACAGAT------ 86001

Qy    282 LeuAlaLeuArgLeuLysSerSerProArgGlyLeuGlnLeuGluLeuGlnAsnAlaGluVal 301
                |||||||:::
Db  86002 ---------------AAAGAAATTAGTGATACAGAAAAGTAAA 86028

Qy    302 GlnSerLeuLeuSerTyrTyrGlnAsnArgMetArgAspProLeuSerLeuSerArgArgProLeu 321
                :::                            |||||:::::
Db  86029 TATACAGAAGCAGGCGAAGATGAGTCAGAGGAAGCACCCCAGAATCCGATGATGCCGGCGATG 86088

Qy    322 AlaMetAsnAsp---GluAspGluSerAlaPheArgAlaMetGluAla 336
                |||          :::||||||
Db  86089 GAATCAGGTGATTCTGACCGAGAGGTGGTAGTGGCACTGGCGCTCCAACGCCGCCTGGT 86148

Qy    337 ---------ArgAlaAlaLysLeuAspArgGlnLysSerGlnLeuValLeuGlyLeuLeuHisGlyPhe 353
                         |||||||||:::
Db  86149 GATGGCATTCGCGAAGACGCCGATGACTTCGATCGGTTATGCGCTCTGGCCCGCTGGCCAT 86208

Qy    354 GlyGluSerAsp--------------------AspAspGluAsp 361
                |||||||||                        :::||||||||
Db  86209 GGGGATAGTGATAACGGCGAAGACTCCGATGCCGGTCGTGAGGAGTCCGAAGGAGAAGAC 86268

Qy    362 GluGluAspGluAsnLeuIleAspProSerGluAsnSerPheArg------------------- 376
                |||          :::|||       |||:::
Db  86269 GAAGGCGAAGAGAGACGTCGCGATCACACTGACGACGGCCAAGGAAAATGAGGATGCC 86328

Qy    377 ---ArgAlaProLeuArgLeuSerSerGlyPheVal--------------GluLysLeuLysSer 392
                   :::                            :::::             |||||:::
Db  86329 GCCGAAGGCGCCTGAACAGAGTCAGGCTTTGTTCCTGCGCCGGATGAAATGACTCTG 86388

Qy    393 AsnAspGluLeuLysSerSerAlaLeuAspArgArgIleLysTyrArgValAspAspVal------ 410
                |||||||||          ::::::          :::::::::
Db  86389 GAGGACGCGCACTCAAG--GCGCTCGACGAGATGGAAGAAGGACAGAAATGACCGAA 86445

Qy    411 --------------GluLysTyrLeuAlaProLysProMetGluPheAsn 422
                              :::::       ||||::::
Db  86446 GACGCACTGTCGCGCCACCATCACAAGAGCTTATGAGCGCGCCTCACTTTTCTGAGTATCGC 86505

Qy    423 ProLysProGlyGlyTyrPheAlaProArgLysIleProThrArgProArgLysMet 442
                |||::::                                    :::
Db  86506 CCATACGGATCGTTCATACGACTTT-------- 86529

Qy    443 LeuProLeuLeuIleLysSerValGlnGluGluIleArgArg---------- 459
                ||||||:::              :::    :::|||
```

Right column

```
Db  86530 --------ATCGGGTTGATTGATGAGCGCGATGAGAGAGCATGTAAAACGCACCAGAAAG 86577
                                                                        |||
Qy    460 --------------HisProSerThrGluTrpLysIleAlaAlaLysGluSer 471
Db  86578 ACAATCGGCGGCAATCCCAATGCACTCCACCTGTCGATCGCTACCGATGGTTCCGGAAGGC 86637

Qy    472 ArgValLeuThrAsnLeuLysAsnAsnProSerLeuAlaAla--------------Leu 486
                                |||                :::||:
Db  86638 AGAAAACTCTTGAACTACAAAAACATGAAAAACATCTGTCTGCCGCCGGGTTCTTCGACTCTG 86697

Qy    487 PheMetAspAspProGlnMetThrLeuLysGlyArg--------GlnMetLeuThr 503
                :::                                    |||
Db  86698 GCCAAAGAC--------CTGGAGCGAGCTATCGCCAGCCGCAACCGAGTTCAGTTCATACCG 86751

Qy    504 AspGluGlnLysGlyArgThrArgValLysThrArgIleThrArgAlaLeuProArgLeuPheGly 523
                                                    |||:::::::::::
Db  86752 GGTCAGAGACGTGGG------------------------CGGATACATGGC 86778

Qy    524 AlaProThrAlaLysAlaGluMetIleAspAlaLysValPheGlnAsnIleGluGlu--- 542
                |||                     :::                   |||:::::::
Db  86779 GCGGAACCCTGTATCGTCTGGCAATGAACGACCCGGTTCCCAAGAAAGAAGACCAC 86838

Qy    543 -------------------------ArgProLeuProPheGluProLysGlyArg 555
                |||                        :::       ::::::
Db  86839 AGAGCGGGTGAACGCGTGCCTGCGTCGATCGTCATGGGCGGTCGTCAATGGGCGCAGG 86898

Qy    556 HisThrArgLeuArgTrpThrGlyTyrAlaAsnGluLysGluIleProGlyLeuGlySerArg 575
                :::::                                           |||
Db  86899 AAGATTCAACTG--------------------GCACTCCGCCAGCGCC 86925

Qy    576 Phe---IleLeuProSerLeuAspPro---ThrMetProAlaLeuAsnThrAlaPheSer 593
                :::       :::||:::       |||||
Db  86926 TACACCATTGCGGATGCTCCTGGATCGTCGTATCGTCGTAACATCATCACCGGCTTTACA 86985

Qy    594 ThrGlnGlyTyrArgAlaArgAspGluTrpAspThrMetPheLys------------- 607
                :::                            |||||
Db  86986 ACGTTTGGT-----AGCCCAGATTATGAAACCATGTTCGAAGCGCGGGGGTTTACACGTTTC 87039

Qy    608 --------------------IleProAsnAsnTrpAsn 613
                                    :::|||
Db  87040 GAGGCGCCTCATGCTGCCCATTATCAAAAACTGGAAT 87075
```

RESULT 10

```
RESULT 10
ADC13546
ID   ADC13546 standard; cDNA; 6126 BP.
XX
AC   ADC13546;
XX
DT   18-DEC-2003  (first entry)
XX
DE   Human NOVX protein encoding cDNA sequence, SEQ ID No 25.
XX
KW   NOVX; FADD interacting protein; ATPase; H+ Transporting; Lysosomal;
KW   FGF 17; Single Pass Transmembrane; Beta-Ketoacyl Synthase; Neuralin 2;
KW   Glutamate Receptor Interacting Protein 2; ChR-Methyltransferase;
KW   NP25 Variant; GTPase-Activating Protein; ELKS; Sim2; RhoGAP;
KW   Phospholipase; Scavenger Receptor Domain Containing Protein;
KW   Metallochionein IA; NOGO receptor; FYVE; NOELIN;
KW   Cyclin Regulatory Subunit; Tetratrico Peptide Repeat Protein;
KW   Immunoglobulin Domain Containing Protein; PA Domain Containing Protein;
KW   Phenylalanine; Histidine Ammonia-Lyase; Cellular Retinaldehyde-Binding;
KW   Glutamine Repeat Containing Protein; TNF Receptor Associated Factor2;
KW   Vacuolar Protein Sorting Homologue R-VPS33A;
KW   Bola Domain Containing Protein; Neurotrophin Receptor;
KW   RAL Guanine Nucleotide Dissociation Stimulator; Armadillo/Beta-Catenin;
KW   Metalloprotease; T10 Ser/Thr-rich; Ring finger-like; cytostatic;
KW   gene therapy; vaccine; cancer; gene; ss.
XX
OS   Homo sapiens.
XX
PN   WO2003004617-A2.
XX
PD   16-JAN-2003.
```

```
Qy   36 LeuValGlyLysSerHisGlnLysProLeuAlaProSerMetGluAlaLeuGluLeu  55
           |||   ||||||
Db  602 CTGACTGGCTCATCCATCATCACCACCAGGTCCCTCCATG---TTGAGGCAGGTA  658

Qy   56 MetGlyValGlnPheValAspAlaLeuIleLysLysGlyGlnMetGlyMetAlaLysGly  75
                                   :::|||
Db  659 AGAGACAGCACAATGTTAGATCTTCAGGCCCAGCTGAAAGAACTGCAGAGAGAATGAC  718

Qy   76 AlaPheLysThrGlnLeuGluValLeuGlnLysLysValHisProAspGlnPheAspLysTyr  95
                                     :::    |||
Db  719 CTCCTCCGGAAAGAGCTAGACATCAAGGACAGCAAATTGGATCTTCCATGAACAGTATT  778

Qy   96 LysLysLeuLysValAspAspLeuAlaAlaAspAlaValMetGlnAlaLysAlaGluMetAla  115
           ||   :::|||    :::    :::||||||||::|
Db  779 AAGACTTTCTGGAGTCCTGAGCTTAAGAGAGGAGAGTCTTGAGGAAGAAGAGGAGCCAGCG  838

Qy  116 LysLeuGlnProLysSerGlyAsnAlaPheIleAspMetLeuAsnGlyAsnGlyIlePro  135
                                :::|||
Db  839 CGGATGTCTGTCTTCAAGGAGCAGAGATGAGGGTTCCATGAAGAAAATCAGCACCTACAG  898

Qy  136 IleGlySerSerIleArgGlyLeuGluAspAlaPheAlaIleArgThrGlnArgAspMetGluAsn  155
            :::||||||   :::|||   :::|||
Db  899 TTG------ACAATCCAGGCCCTTCAAGATGAGCTGCGAACCCAGAGAGACCTCAACCAC  952

Qy  156 ThrAspProSerGluGlnIleIleAlaGlyAlaValMetAspLysPheGlnThrGlnIleLeu  175
            |||
Db  953 CTCCTCCAGCAAGAGAGTGGCAACCAGCAGCG---GAGCACTTCACCATGCAGCTGACC  1009

Qy  176 ProGlyLeuValAlaAsnMetIleAlaGlyLysAsnProPheLysMetProGlnGlnMet  195
                   :::    |||
Db 1010 ------------------------GAGGAGAACTTTAGGCGGCTCCAAGCCGAG  1039

Qy  196 ArgLysAlaGlnAlaAlaProSerSerValPheGlnGlnAlaLeuAlaAlaGlnAlaArgAlaMet  215
            :::    :::    :::    |||
Db 1040 CATGAGACAGGCCTAAGGAGCTGTTCCTTTGAGGAAGACATTAGGAAGAAATGGAGCTG  1099

Qy  216 LeuGlyLysAsnAlaProValAlaGlyAspAlaGlyGlyGlnAsnArgMetMetMetAsn  235
                                   :::|||
Db 1100 ------------------AGAATTGAAACGCAGAAACAAACCCTCAAT  1129

Qy  236 ArgValAspArgMetGlnGln------ArgGluLeuGlnGlnLeuAspGluAspAsp  253
            :::|||   :::||||||    :::
Db 1130 GCCCGAGATGCAATTAAAAAACTTCTTGAGATGTTGCAAAGTAAAGGCTTGCCATCC  1189

Qy  254 AspAspLeuGlnAspGluAspValProArgArgArgSerSerAspGlyGluProGln--  272
            ||||||   :::
Db 1190 AAAAGCCTGAGGATGACAATGAGCCAGACGGCGATGAGGCTCAGGTCTCAGGTC  1249

Qy  273 ------------------SerGluAlaGluHisGlnArgArgArgAsp  281
                                 :::
Db 1250 AGCCACTTGGAAGTGATTTTAGATCAGAAGAAGAAGAAAACATACATCTTAGAGAGGAA  1309

Qy  282 LeuAlaArgLeuLysSerSerProArgLeuLeuGlnLeuAsnAlaGlnVal  301
            |||   :::
Db 1310 TTGCACCGAAGAGAGCCAACTTCAGCCG------------GAGCCAGCCAAGACG  1351

Qy  302 GlnSerLeuLeuSerTyrGlnArgMetArgAspSerProLeuSerLysArgProLeu  321
            |||:::||:::||
Db 1352 AAGGCTCTCCAGACTGTCATCGAAATGAAGGACACAAAAATCGCTTCA------TTG  1402

Qy  322 AlaMetAsnAspGluAsp--GluSerAlaPheArgGluAlaMetGluAlaLysLeu  340
            |||   ||||::|||
Db 1403 GAACGAAACATAAGGATCTTGAGGATGAGATCCAGATGTTAAAAGCCAATGTGTGTGTCTG  1462

Qy  341 AspGlnLysSerGlnLeuValLeuHisGlyPheGlyGluSerAspAspArgGlu  360
            ||
Db 1463 AACACTCAG  1471

Qy  361 AspGluAspGluAsnLeuIleAspProSerGluAsnSerPheArgArgAlaProLeu  380
            |||:::||:::|||
Db 1472 GACCCGGCCAAGAGAGAG-  1486

Qy  381 ArgLeuSerSerGlyPheValGluLysLeuSerAsnAspGluLeuLysSerAlaLeu  400
```

XX
PF 03-JUL-2002; 2002WO-US021359.
XX
PR 05-JUL-2001; 2001US-0303046P.
PR 09-JUL-2001; 2001US-0303828P.
PR 11-JUL-2001; 2001US-0304502P.
PR 12-JUL-2001; 2001US-0305011P.
PR 13-JUL-2001; 2001US-0305262P.
PR 17-JUL-2001; 2001US-0306085P.
PR 24-JUL-2001; 2001US-0307536P.
PR 27-JUL-2001; 2001US-0308228P.
PR 30-JUL-2001; 2001US-0308877P.
PR 01-AUG-2001; 2001US-0309255P.
PR 10-AUG-2001; 2001US-0311753P.
PR 19-SEP-2001; 2001US-0323449P.
PR 22-FEB-2002; 2002US-0358932P.
PR 05-MAR-2002; 2002US-0361765P.
XX
XX
PA (CURA-) CURAGEN CORP.
XX
PI Patturajan M, Gerlach VL, Anderson DW, Taupier RJ, Zerhusen BD;
PI Guo X, Casman SJ, Hjalt T, Miller CE, Kekuda R, Shimkets RA;
PI Malyankar UM, Zhong M, Padigaru M, Li L, Shenoy SG, Gorman L;
PI Edinger SR;
XX
DR WPI; 2003-201550/19.
DR P-PSDB; ADC13547.
XX
PT New NOVX polypeptide, useful for preparing a composition for treating or
PT preventing cancer.
XX
PS Claim 20; Page 116-117; 393pp; English.
XX
CC The invention relates to a novel isolated NOVX polypeptide comprising; a
CC sequence of 57-1149 amino acids as defined in the specification, or its
CC mature form; a sequence that is at least 95% identical to the 57-1149
CC amino acid polypeptide; or a sequence comprising one or more conservative
CC substitutions in the 57-1149 amino acid polypeptide. The NOVX proteins of
CC the invention include the following protein families: FADD interacting
CC protein-like, ATPase, H+ Transporting, Lysosomal (vacuolar Proton Pump)-
CC like, FGF 17-like, Single Pass Transmembrane-like, Beta-Ketoacyl Synthase
CC -like, Neuralin 2-like, Glutamate Receptor Interacting Protein 2-like,
CC ChR-Methyltransferase-like, NP25 Variant-like, GTPase-Activating Protein-
CC like, ELKS-like, Sim2-like, RhoGAP-like, Phospholipase-like, Scavenger
CC Receptor Domain Containing Protein-like, Metallothionein IA-like, NOGO
CC receptor-like, FYVE-protein, NOELIN-like, Cyclin Regulatory Subunit-like,
CC Tetratrico Peptide Repeat Protein-like, Immunoglobulin Domain Containing
CC Protein-like, PA Domain Containing Protein-like, Phenylalanine and
CC Histidine Ammonia-Lyase-like, Cellular Retinaldehyde-Binding-like,
CC Glutamine Repeat Containing Protein-like, TNF Receptor Associated Factor2
CC -like, Vacuolar Protein Sorting Homologue R-VPS33A, Bola Domain
CC Containing Protein-like, Neurotrophin Receptor-like, RAL Guanine
CC Nucleotide Dissociation Stimulator-like, Armadillo/Beta-Catenin-like,
CC Metalloprotease-like, T10 Ser/Thr-rich-like, and Ring finger-like
CC protein. The NOVX proteins and the encoding polynucleotides have
CC cytostatic activity and can be used in gene therapy or a vaccine. The
CC NOVX polypeptide is useful for preparing a composition for treating or
CC preventing cancer. This polynucleotide sequence represents a gene
CC encoding a NOVX protein of the invention.
XX
SQ Sequence 6126 BP; 1943 A; 1234 C; 1384 G; 1565 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:   0.00176   Length: 6126
Score: 169.50   Matches: 118
Percent Similarity: 36.28%   Conservative: 108
Best Local Similarity: 18.94%   Mismatches: 248
Query Match: 5.15%   Indels: 149
DB: 10   Gaps: 22

US-10-736-868-2 (1-643) x ADC13546 (1-6126)

```
Db   1487 ----------------------------ATCAAACAAATTGAGGTTTACAAAAGTCACTCC 1519
                                      :::|||        |||||||
Qy    401 AspArgIleLysTyrArgValAspAspValGluLysTyrLeuAlaProLysProMetGlu 420
          |||  :::   ||||   :::::::     |||     |||     |||
Db   1520 AAGTTTATGAAGACCAAGATTGATCAGCTGAAGACTTTCGAAGAGAGAGTCGGAA 1579

Qy    421 PheAsnProLysProGlnProGlyTyrPheAlaProArgLysIleProThrArgProArg 440
                                  |||:::   |||:::  |||
Db   1580 CTT-------------------CTTGCCTTACAAACAAGCTTGAAACCCTCAGCAAT 1618

Qy    441 LysMetLeuProLeuLeuIleGlySerAspProLysValGlnGluGlyIleArgArgHis 460
          :::                      |||||   |||       |||:::  :::
Db   1619 CAA------------------AATTCAGATTGCAAGCAACACATTGAAGTGCTCAAAGAG 1660

Qy    461 ProSerThrGluTrpLysIleLysSerArgValLeuThrAsnLeuLysAsnLysAsnAsn 480
          |||                      ||||||||  |||
Db   1661 TCACTTACT-------------GCCAAAGACACAGAGGGCTGCCATC---CTTCAGACTGAG 1705

Qy    481 ProSerLeuAlaLeuPheMetAspAspLys--------LeuGluAsnThrLeuLys 497
          |||              |||  :::::::|||
Db   1706 GTAGATGCGTGAGATTACGACTGGAAGAAAAAGAATCTTCCTCAATAAAAAAACAAA 1765

Qy    498 GlyArgGlnMetLeuThrAspGluGlnLysGlyArgThrArgValLysThrIleArgArgAla 517
          |||||||||||:::||||||
Db   1766 CAGCTACAGGACCTCACA--GAAGAGAAGGGG 1795

Qy    518 LeuProArgLeuPheGlyAlaProThrAlaLysAlaGluMetIleAspAlaAlaLysValPhe 537
          ||||||          |||:::          |||  |||
Db   1796 ---------------ACACTGCGCCGGTGAAATTCGTGACATGAAAGAGATATG 1831

Qy    538 GlnAspIleGluArgProIleProLeuPhePheGluProLysProLysArgHisThr 557
          :::::::|||||||  |||          :::      |||
Db   1832 TTAGAAGTGAAGAAGAAAAAATCAATGTTCTTCAGAAAAAGAATTGAAAACTTGCAAGAA 1891

Qy    558 ArgLeuArgTrpThrGlyValAlaAsnGluLysGlyIleLeuProGlyLeuGlySerArgPheIle 577
          :::|||||||   :::||||||        |||
Db   1892 CAACTTAGG------GATAAAGACAACTGACCAACCTGAAAGACAGA------ 1936

Qy    578 LeuProSerLeuAspProThrMetProAlaLeuAsnThrAlaPheSerThr------ 594
          :::   ||||||           ::|||||||     ::|||
Db   1937 GTGAAGTCCTTGCGACGATTCCAGTATACAGATACTGCACGCGCTAGAGAA 1996

Qy    595 -----------------------GlnGlyArgAlaArgAspGlu 601
                                  :::  ||||||   |||:::
Db   1997 GCTCTGTCAGAGAAGGAGGAGAATAATTGAGCGCTTGAAAGAACAGCGAGAAAGAGATGAT 2056

Qy    602 TrpAspThrMetPheLysIleProAsnAsnTrpAsnProGlyLysAspGluValGlyPheLys 621
          :::  :::||||||             :::::::
Db   2057 CGGGAAAGACTAGAAGAGATAGAATCCTTCCGAAAAGAGAACAAAGACCTGAAAGAGAAG 2116

Qy    622 MetAsnSer 624
          :::||||:::
Db   2117 GTCAATGCT 2125
```

```
RESULT 11
ABZ11646
ID   ABZ11646 standard; cDNA; 6151 BP.
XX
AC   ABZ11646;
XX
DT   20-JAN-2003  (first entry)
XX
DE   Human polynucleotide SEQ ID NO 528.
XX
KW   Human; genome mapping; gene therapy; food supplement; virus; fungus;
KW   cell-proliferative disorder; neurodegenerative disease; bacterial;
KW   Parkinson's disease; Alzheimer's disease; autoimmune disease;
KW   multiple sclerosis; diabetes; genetic disorder; wound; burn; infection;
KW   arthritis; cytostatic; immunomodulator; nootropic; neuroprotective;
KW   antiparkinsonian; antidiabetic; immunosuppressive; dermatological;
KW   haemostatic; vulnerary; fungicide; antibacterial; virucide; protozoacide;
KW   antiarthritic; gene; ss.
```

```
XX   OS   Homo sapiens.
XX
PN   WO20270539-A2.
XX
PD   12-SEP-2002.
XX
PF   05-MAR-2002; 2002WO-US005095.
XX
PR   05-MAR-2001; 2001US-00799451.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT, Zhou P, Goodrich RW, Asundi V, Zhang J, Zhao QA, Ren F;
PI   Xue AJ, Yang Y, Ma Y, Yamazaki V, Chen R, Wang Z, Ghosh M;
PI   Wehrman T, Wang J, Wang D, Drmanac RT;
XX
DR   WPI; 2002-759812/82.
DR   P-PSDB; ABP69429.
XX
PT   New polynucleotides comprising sequences assembled from expressed
PT   sequence tags (ESTs), useful for treating cell-proliferative,
PT   neurodegenerative, autoimmune, genetic, myeloid or lymphoid, or platelet
PT   or coagulation disorders.
XX
PS   Claim 1; SEQ ID NO 528; 1012pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated polynucleotide (I) comprising a
CC   nucleotide sequence selected from any of 948 sequences (ABZ11119-
CC   ABZ12066) or their mature protein coding portion, active domain coding
CC   protein or complementary sequences. The polynucleotides are useful for
CC   identifying expressed genes or for physical mapping of human genome. The
CC   encoded polypeptides (ABP68902-ABP69849) are useful as molecular weight
CC   markers, as a food supplement, for generating antibodies, in medical
CC   imaging, screening and diagnostic assays and for treating cell-
CC   proliferative disorders (cancer), neurodegenerative diseases (Parkinson's
CC   or Alzheimer's disease), autoimmune diseases (multiple sclerosis,
CC   diabetes, lupus) genetic disorders, myeloid or lymphoid disorders,
CC   platelet or coagulation disorders, wound, burns, incision, ulcers, liver
CC   or lung fibrosis, infections (bacterial, viral, fungal, parasitic),
CC   arthritis, etc. Note: The sequence data for this patent did not form part
CC   of the printed specification, but was obtained in electronic format
CC   directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 6151 BP; 1939 A; 1247 C; 1394 G; 1571 T; 0 U; 0 Other;
```

```
Alignment Scores:
Pred. No.:                                Length:            6151
Score:               0.00177              Matches:           118
Percent Similarity:  36.28%               Conservative:      108
Best Local Similarity: 18.94%             Mismatches:        248
Query Match:                              Indels:            149
DB:                                       Gaps:              22

Sequence 6151 BP; 1939 A; 1247 C; 1394 G; 1571 T; 0 U; 0 Other;
```

```
US-10-736-868-2 (1-643) x ABZ11646 (1-6151)

Qy     36 LeuValGlyLysSerHisGlnLysLeuProLeuAlaProSerMetGluAlaLeuGluLeu 55
          |||  |||           |||||||        |||||||
Db    632 CTGACTGGCTCATCCCATCATCACCACCACCAGGTCCCCTCATG---TTGAGGCAGGTA 688

Qy     56 MetGlyValGlnPheValAspAlaLeuIleLeuLysGlyGlnMetAlaLysGlyGly 75
                   :::            |||               :::
Db    689 AGAGACAGCACACAATGTTAGATCTTCAGGCCCCAGCTGAAGAAGAACTGCAGAGAATGAC 748

Qy     76 AlaPheLysThrGlnLeuGluValLeuGluLysValHisProAspPheAspLysTyr 95
                                 |||                  |||:::
Db    749 CTCCTCCGAAAGAGCTAGACACATCAAGGACAGCAAATTGGGATCTTCCATGAACAGTATT 808

Qy     96 LysLysLeuLysValAspAspLeuAlaAlaAspAlaValMetGlnIleAlaGluMetAla 115
                 |||                :::|||     |||
Db    809 AAGACTTTCTGAGTCTGAGTTAAGAAGGAGAGTCTTGAGGAAAGAAGAGGCAGCG 868

Qy    116 LysLeuGlnProLysSerGlyAsnAlaPheIleAspMetLeuAsnGlyAsnGlyIleLysPro 135
```

Qy   869 CGGATGTCTGTCCTCAAGGAGCAGATGAGGGTTTCCCATGAAGAAATCAGCACCTACAG 928
              :::::|:   :::
Db   136 IleGlySerIleArgGlyLeuGluAspAlaIleArgThrGlnArgMetGluAsn   155
              ::: |||||:||
Qy   929 TTG----ACAATCCAGGGCCTAAGGAGCTGTCGAACCCAGAGAGACCTCAACCAC   982

Db   156 ThrAspProSerGluGlnIleAlaLysAlaValMetAspLysPheGlnThrGlnIleLeu 175
              ||| :::   ::::::
Qy   983 CTCCTCCAGCAGAGAGTGGCAACCGAGACGCG---GAGCACTTCACCATCGAGCTGACC 1039

Db   176 ProGlyLeuValAlaAsnMetIleAlaGlyLysAspPheLysMetProGlnGlnMet   195
              |||
Qy   1040 ----------------GAGGAGAACTTTAGGCGGCGCTCCAAGCCGAG   1069

Db   196 ArgLysAlaGlnAlaAlaProSerSerValPheGlnIleAlaLeuAlaAlaGlnArgAlaMet 215
              ||||| :::   :::::
Qy   1070 CATGACAGGCCAGGCTAAGGAGCTGTCCTTTGAGGAAGACATTAGAGGAAATGGAGCTG 1129

Db   216 LeuGlyLysAsnAlaProValAlaGlyGlyGlyValGlyArgMetMetMetAsn   235
              ||||| :::||||
Qy   1130 --------------AGAATTGAAACGCAGAAACAAACCCTCAAT   1159

Db   236 ArgValAspArgMetGlnGln------ArgGluLeuGlnGluGluAspGluAspAsp   253
              :|| ::::::|||
Qy   1160 GCCCGAGATGAGTCAATTAAAAAAACTTCTGAGATGTGCAAAGTAAAGGCTTGCCATCC 1219

Db   254 AspAspLeuGluAspValProArgArgArgSerSerArgAspGlyGluProGln---   272
              ||||||:::
Qy   1220 AAAAGGCCTGGAGGATGACAATGACGACGCGGCCGACGGCCGATGGCAGAGGCTGAGTCCCAGGTC 1279

Db   273 ------SerGluAlaGluHisGlnIleArgArgAsp   281
                    ||| :::
Qy   1280 AGCCACTTGGAAGTGATTTTAGATCGAAAGAAGAACATTACATCTTAGAGAGGAA   1339

Db   282 LeuAlaAlaArgArgLeuLysSerSerProArgLeuLysGluLysLeuGluGluLeuGlnAsnAlaGluGluVal 301
              ||| :::
Qy   1340 TTTGCACCGGAAGAAGCCAACTTCAGCCG----------------GAGCCAGGCCAAGACACG 1381

Db   302 GlnSerLeuLeuSerTyrGlnArgMetArgAspSerProLeuSerArgArgProLeu   321
              :::::|||| :::
Qy   1382 AAGGCTCTCCAGACTGTCATCGAAATGAAGGACACAAAATCGCTTCA---------TTG 1432

Db   322 AlaMetAsnAspGluAsp---GluSerAlaPheArgAlaMetGluAlaArgAlaLysLeu 340
              ||| :::
Qy   1433 GAACGAAACATAAGGGGATCTTGAGGATGAGATCCAATGGTGTGCTG   1492

Db   341 AspGlnLysSerGlnLeuValLeuGlyLeuGlyLeuHisGlyPheGlyGluSerAspAspGlu 360
              :::   :::
Qy   1493 ----------------AACACTGAG   1501

Db   361 AspGluGlyAspAspAsnLeuIleAspProSerGluAsnSerPheArgArgArgAlaProLeu 380
              |||||||
Qy   1502 GACCGCGGAAGAACAG   1516

Db   381 ArgLeuSerSerGlyPheValGluLeuLysSerAsnAspGluLeuLysSerAlaLeu   400
              ::|||
Qy   1517 ----------------ATCAACAAATTGAGGTTTACAAAGTCACTCC   1549

Db   401 AspArgIleLysTyrArgValAspAspGlnLysSerTyrLeuAlaLysProLysProMetGlu 420
              :::::|||| :::
Qy   1550 AAGTTTATGAAGAACCAAGATTGATCAGCTGAAGCAGGAACTTTCAAAGAAAGAGTCGGAA 1609

Db   421 PheAsnProLysProGlnProGlyTyrPheAlaProArgLysIleProThrArgProArg 440
              ||| :::
Qy   1610 CTT----CTTGCCTTACAAACAAGCTTGAAACCCTTGAAAACCCTCAGCAAT   1648

Db   441 LysMetLeuProLeuLeuIleGlyGluLeuIleGlySerAspProLysValGlnGluGluIleArgArgHis 460
              ||||| :::   :::
Qy   1649 CAA----AATTCAGATTGCAAGCAACAACATTGAAGTGCTCAAAGAG   1690

Db   461 ProSerThrGluTrpLysIleAlaLysGlySerArgValSerArgValLeuThrAsnLeuLysAsnAsn 480
              ||||||| |||:::

Db   1691 TCACTTACT--------GCCAAAGACACAGAGGGCTGCCATC---CTTCAGACTGAG   1735
              ::::::::|||
Qy   481 ProSerLeuAlaAlaLeuPheMetAspAspLys------LeuGluAsnThrLeuLys   497
              ::: |||
Db   1736 GTAGATGCGTGAGATTACGACTGGAGAAAAAAGAATCTTTCCTCAATAAAAAAACAAAA 1795

Qy   498 GlyArgGlnMetLeuThrAspArgAspGluGlnLysLysGlyArgThrArgValLysThrIleArgArgAla 517
              ||||| :::::||||
Db   1796 CAGCTACCAGGACCTCACA--GAAGAGAAGGGG   1825

Qy   518 LeuProArgLeuPheGlyAlaProThrAlaLysAlaGluMetIleAspAlaLysAlaValPhe 537
              ||| :::
Db   1826 ------------ACACTGGCGGTGAAATTCGTGACATGAAAGATATG   1861

Qy   538 GlnAspIleGluGlnArgProIleProProLeuProPheGluProLysGlyArgHisThr   557
              :::::::::||||| :::
Db   1862 TTAGAAGTGAAGGAAAGAAAAATCAATGTTCTTCAGAAAAAGATTGAAAACTTGCAAGAA 1921

Qy   558 ArgLeuArgTrpThrGlyAlaAsnGluLeuProGlyLeuGlySerArgPheIle------ 577
              ::::|||||
Db   1922 CAACTTAGG----GATAAAGACAAGCAACTGACAGAAACAGACAGA------------- 1966

Qy   578 LeuProSerLeuAspProThrMetProAlaLeuAsnThrLeuPheSerThr-------- 594
              ::||||||| :::||||
Db   1967 GTGAAGTCCTTGCAGACGGATTCCAGTAATACAGATACTGCCACTGCGCCGACGTGAGGAGAA 2026

Qy   595 --------------------------------GlnGlyValArgGlyAlaIleArgArgAspAspArgGlu 601
              :::|||||
Db   2027 GCTCTCGTCAGAGAAGAGAATAATTGAGCGCGTTGAAAGAACACGCGAGAAAGAGAATGAT 2086

Qy   602 TrpAspThrMetPheLysIleProAlaAsnAsnTrpAsnProGlyAspGluValGlyPheLys 621
              ::: :::||||
Db   2087 CGGGAAGAACTGAAGAGATAGAAATCCTTCCGAAAAGAGAAGAGAACAAAGACCTGAAAGAGAAG 2146

Qy   622 MetAsnSer   624
              :::||||||
Db   2147 GTCAATGCT   2155

RESULT 12
ADM44164
ID   ADM44164 standard; cDNA; 6151 BP.
XX
AC   ADM44164;
XX
DT   03-JUN-2004 (first entry)
XX
DE   Novel human arginine-rich protein cDNA #528.
XX
KW   ss; gene; human; arginine-rich protein; cancer; inflammation;
KW   genetic disorder.
XX
OS   Homo sapiens.
XX
PN   US2004053250-A1.
XX
PD   18-MAR-2004.
XX
PF   21-NOV-2002; 2002US-00302172.
XX
PR   05-MAR-2001; 2001US-00799451.
PR   05-MAR-2002; 2002WO-US005095.
PR   20-AUG-2002; 2002US-00225251.
XX
PA   (TANG/) TANG Y T.
PA   (XUEA/) XUE A.
PA   (DRMA/) DRMANAC R T.
XX
PI   Tang YT, Xue A, Drmanac RT;
XX
DR   WPI; 2004-238579/22.
XX
PT   New isolated arginine-rich protein-like polynucleotides and polypeptides,
PT   useful for diagnosing and/or treating conditions associated with aberrant

PT activity of the arginine-rich polypeptides, such as cancer and
PT inflammation.
XX
PS Disclosure; SEQ ID NO 528; 51pp; English.
XX
CC The invention relates to an isolated polynucleotide. The methods and/or
CC compositions of the present invention are useful for the diagnosis and/or
CC treatment of diseases or conditions associated with aberrant expression
CC or activity of the arginine-rich protein-like polypeptides, such as
CC cancer and inflammation. They can also be used in forensics, gene
CC mapping, identification of mutations responsible for genetic disorders,
CC and in assessing biodiversity. The present sequence represents a novel
CC human arginine-rich protein cDNA.
XX
SQ Sequence 6151 BP; 1939 A; 1247 C; 1394 G; 1571 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	0.00177	Length:	6151
Score:	169.50	Matches:	118
Percent Similarity:	36.28%	Conservative:	108
Best Local Similarity:	18.94%	Mismatches:	248
Query Match:	5.15%	Indels:	149
DB:		Gaps:	22

US-10-736-868-2 (1-643) x ADM44164 (1-6151)

```
Qy    36 LeuValGlyLysSerHisGlnLysLeuProLeuAlaProSerMetGluAlaLeuGluLeu    55
            |||  ||| |||||              :::::::
Db   632 CTGACTGGTCATCCATCACCACCACCAGTCCCCTCCATG---TTGAGGCAGTA        688

Qy    56 MetGlyValGlnPheValAspAlaLeuIleLysLysGlyGlnMetGluMetAlaLysGly    75
                           :::|||            :::::::::
Db   689 AGAGACAGCACAATGTTAGATCTTCAGGCCCAGCTGAAGAACTGCAGAGAGAATGAC     748

Qy    76 AlaPheLysThrGlnLeuGluValLeuGluValValHisProAspAspGlnPheAspLysTyr    95
            :::|||:||                          :::
Db   749 CTCTCCGGAAAGAGCTAGACATCAAGGACAGCAAATTGGATCTTCCATGAACAGTATT    808

Qy    96 LysLysLeuLysValAspAspPheLeuAlaAspLeuAlaAspAlaAlaMetGlnAlaAlaMetAla   115
            |||               ::  |||          :::::::::::
Db   809 AAGCACTTTCTGGAGTCCTGAGCTTAAGACAGGAGAGAGTCTTGAGGAAAGAAGAGGCAGCG   868

Qy   116 LysLeuGlnProLysSerGlyAsnAlaPheLeuAspMetLeuAsnGlyAsnGlyIleIlePro   135
            :::
Db   869 CGGATGTCTGTCCTCAAGGACGAGATGAGGGGTTTCCCATGAAGAAAATCAGCACCTACAG   928

Qy   136 IleGlySerSerIleAlaGlyLeuGluAspGluAlaIleArgThrGlnArgAspMetGluAsn   155
            ::  |||:|||   |||:||
Db   929 TTG-----ACAATCCAGGCCCTTCAAGATGAGCTGCCAACCCAGAGAGACCTCAACCAC    982

Qy   156 ThrAspProSerGluGlnIleIleAlaLysAlaValMetAspLysPheAspLysHisThrGlnThrIleLeu  175
            :::   :::
Db   983 CTCTCCAGCAAGAGAGTGCAACCGACGGAGCG--GAGCACTCACCATCGAGCTGACC   1039

Qy   176 ProGlyLeuValAlaAsnMetIleAlaGlyLysLysAsnProPheLysMetProGlyGlnMet   195
            |||:::
Db  1040 --------------GAGGAGAACTTTAGGCGGCTCCAAGCCGAG                1069

Qy   196 ArgLysAlaGlnAlaAlaProSerSerValPheGlnAlaAlaLeuAlaLeuAlaGlnAlaArgAlaMet  215
            ||||||
Db  1070 CATGACACAGGCAGGCTAAGGAGCTTGTCCTTTGGGAAGACAGATTAGAGGAAATGGGAGCTG  1129

Qy   216 LeuGlyLysAsnAlaProValAlaGlyValAlaProValAlaGlyGluGluGlyArgMetMetMetAsn   235
            |||   |||:|||
Db  1130 AGAATTGAAACGCCAGAAACAAACCCTCAAT                             1159

Qy   236 ArgValAspArgMetGln------ArgGlyLeuGlyGlnGluGlnGluAspAspAsp   253
            ||||:  :::::::
Db  1160 GCCCGAGATGAGTCAATTAAAAACTTCTGAGATGTTGCAAAGTAAAGGCCTTGCCATCC  1219

Qy   254 AspAspLeuGluAspGluLeuGlnValProArgArgSerSerAspGlySerGlyProGln---  272
            :::   :::::::
Db  1220 AAAAGCCTGGAGGATGACAATGACGGACGCCGGCGAACGCGCGGATGGCAGGCTGAGTCTCAGTC  1279
```

```
Qy   273 ---------------------------------SerGluAlaGluHisGlnArgArgAsp   281
Db  1280 AGCCACTTGGAAGTCGATTTAGATCAGAAAGAAGAAGGAGAAACATACATCTTAGAGAGGAA  1339

Qy   282 LeuAlaArgArgLeuLysSerSerProArgLeuLysGluLeuLeuLysGlnLeuAsnAlaGluVal  301
            |||                          :::  |||||:|||:
Db  1340 TTGCACCGAGGAAGCCAACTTCAGCG--------------GAGCCAGCCAAGACG         1381

Qy   302 GlnSerLeuLeuSerTyrGlnArgMetArgAspSerProLeuSerArgArgProLeu   321
            |||:||:||   ::::::
Db  1382 AAGGCTCTCCAGACTGTCATCGAAATGAAGGACACAAAAATCGCTTCA------TTG    1432

Qy   322 AlaMetAsnAspGluAsp--GluSerAlaPheAlaMetGluAlaArgAlaLysLeu   340
            :::
Db  1433 GAACGAAACATAAGGGATCTTGAGGATGAGATCCAGATGTTAAAAGCCAATGGTGTGCTG   1492

Qy   341 AspGlnLysSerGlnLeuValLeuGlyPheGlyLeuHisGlyPheGlyGluSerAspAspAspGlu  360
            :::
Db  1493 --------------AACACTGAG                                       1501

Qy   361 AspGluAspAspAsnLeuIleLeuAspProSerPheArgArgAlaProLeu   380
            |||  :::||||
Db  1502 GACCGCGAAGAAGAG                                              1516

Qy   381 ArgLeuSerSerGlyPheValGluLeuCysLysSerAsnAspProLeuLeuLysSerAlaLeu   400
Db  1517 -------------ATCAAACAAATTGAGGTTTACAAAGTCACTCC               1549

Qy   401 AspArgIleLysTyrArgArgAlaAspValGluLysTyrLeuAlaProLysProMetGlu   420
            ::::::|||
Db  1550 AAGTTTATGAAGACCAAGATTGATCAGCTGAAGCAGGAGACTTTCAAAGAAGAGTCGGAA   1609

Qy   421 PheAsnProLysProGlnProGlyTyrPheAlaProArgIleProThrArgProArg   440
Db  1610 CTT--------CTTGCCTTACAAACAGCTTGAAACCCTCAGCAAT               1648

Qy   441 LysMetLeuProLeuLeuLeuSerAspProLysValGlnIleGluIleArgArgHis   460
            |||
Db  1649 CAA----------AATTCAGATTGCAGCAACACATTGAAGTGCTCAAAGAG          1690

Qy   461 ProSerThrGluTrpLysIleAlaLysGluSerArgValLeuThrAsnLeuLysAsnAsn   480
            |||||||
Db  1691 TCACTTACT--------GCCAAAGAACAGAGGGCTGCCATC---CTTCAGACTGAG      1735

Qy   481 ProSerLeuAlaLeuPheMetAspAspLys------------LeuGluAsnThrLeuLys   497
            |||  :::::::::
Db  1736 GTAGATGCCTGAGATTACGACTGGAAGAAAAAGAAATCTTCCTCAATAAAAAAACAAAA   1795

Qy   498 GlyArgGlnMetLeuThrArgAspThrAlaLysThrIleArgArgValLysThrIleArgAla  517
            |||||||||
Db  1796 CAGCTACAGGACCTCACA--GAAGAGAGGGG                             1825

Qy   518 LeuProArgLeuPheGlyAlaProThrAlaLaLysAlaGluMetIleAspAlaLaLysValPhe   537
            |||
Db  1826 --------ACACTGGCCGGTGAATTCGTGACATGAAAGATATG                  1861

Qy   538 GlnAspIleGluArgProIleProProLeuPhePheGluProLysGlyArgGlyHisThr   557
            :::   :::
Db  1862 TTGAAGTGAAGGAAGAAAAATCAATGTTCTTCAGAAAAAGATTGAAAACTTGCAAGA     1921

Qy   558 ArgLeuArgTrpThrGlyValAlaAsnGluLysGluProGlyLeuGlySerArgPheIle   577
            :::|||:||:
Db  1922 CAACTTAGG---------GATAAAGACAGCAACTGACCAACCTGAAAGACAGA--       1966

Qy   578 LeuProSerLeuAspProThrMetProAlaLeuAsnThrAlaPheSerThr-------   594
Db  1967 GTGAAGTCCTTGCAGACCGATTCCAGTAATACAGATACTGCACTGGCACCGCTAGAGGAA  2026

Qy   595 -----------------GlnGlyArgAlaArgAspGlu                      601
            :::   :::
Db  2027 GCTCTGTCCAGAAGGAGAGGAGGAGAATAATTGAGCCGTTGAAAAGAACAGCGAGAAAGAGATGAT  2086
```

```
Qy   602  TrpAspThrMetPheLysIleProAsnAsnTrpAsnProGlyAspGluValGlyPheLys  621
               ::::||||||   ::||||||    ::
Db  2087  CGGGAAAGACTAGAAGAGATAGAATCCTTCCGAAAAGAGACAAAGACCTGAAAGAGAAG  2146

Qy   622  MetAsnSer  624
               :::||||||:::
Db  2147  GTCAATGCT  2155


RESULT 13
ABL02411
ID   ABL02411 standard; cDNA; 5439 BP.
AC   ABL02411;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster expressed polynucleotide SEQ ID NO 1715.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical; gene; ss.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US009231.
XX
PR   23-MAR-2000; 2000US-0191637P.
PR   11-JUL-2000; 2000US-0611450.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC, Adams M, Li PWD, Myers EW;
XX
DR   WPI; 2001-656860/75.
DR   P-PSDB; ABB58308.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signaling and cell-cell
PT   interactions.
XX
PS   Claim 1; SEQ ID NO 1715; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
CC   ABB72072). The sequence data for this patent did not form part of the
CC   printed specification, but was obtained in electronic format directly
CC   from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 5439 BP; 1803 A; 1228 C; 1463 G; 945 T; 0 U; 0 Other;


Alignment Scores:
Pred. No.:              0.00201           Length:         5439
Score:                168.00             Matches:         124
Percent Similarity:    36.36%            Conservative:    100
Best Local Similarity:  5.10%            Mismatches:      230
Query Match:               4             Indels:          162
                                         Gaps:             29
                                         DB:

US-10-736-868-2 (1-643) x ABL02411 (1-5439)

Qy    65  IleLysLys------GlyGlnMetGluMetAlaAlaLysGlyAlaPheLysThrGlnLeuGlu   82
               |||       |||      |||:::||
Db  1102  CTTAAGAAATGAGCAGTAGAAATGAGCTGGAAATGGGCAGCGATAAGGAAGACCAGGAA   1161

Qy    83  ValLeuGluLysValHisProAspLysGlnPheAspLysTyrLysLysLeuLysValAspAsp  102
               |||:::
```


```
Qy        ...                                          :::  |||          |||
Db  1162  ATGCAGGAA---------CCTCCCAAAAAGGCAAAGAAAAATAAACAACGAATGGACAGT  1212

Qy   103  LeuAlaAlaAspAlaValMetGlnIleAlaGluMetAlaLysLeuGlnProLysSerGly  122
               :::            |||        |||:::
Db  1213  GAAAGGCAAGATGAAATC--CCGAAAACCGAGAAGGCGAGAAATAACCAGTTCACCCAAA  1269

Qy   123  AsnAlaPheIleAspMetLeuAsnGlyAsnGly---------  133
               :::||||||
Db  1270  AACAAACTGAAAGGATGGTCGACTCGAATCTGACCGGAGAAACTGCTGAGGAAGTT  1329

Qy   134  IleProIleGlySerSerIleArgGlyLeuGluAspAlaIleAlaIleArgThrGlnArgAspMet  153
               |||:::   |||      |||:::        |||
Db  1330  GACAAATCCCGAGGAATCTGCAGGAGACAGGAAGCCCCAATGGAAAGCGCTCTCCAGA  1386

Qy   154  GluAsnThrAspProSerGluGlnIleIleAlaLysAlaValMetAspLysPhe---------  170
               :::        |||        |||:::
Db  1387  AGTCCGTCAAGAACAAACTTAAAGATTGGTGGACTCG---GAATCGGAACCAGAGCTA  1446

Qy   171  ---------GlnThrGlnIleLeuProGlyLeuValAlaAsnMetIleAlaGlyLys  186
                          :::   |||
Db  1447  GAGAAACCTAAGAAGGCCAAAGTTGTTCGCGAATCAGCTAAGAAAGCACTGAGGGGATG  1506

Qy   187  AsnProPheLysMetProGlnMet------ArgLysAla---------  198
               :::||||||                :::
Db  1507  CAGGCCATCCAAAGTGAACAGCGTTTGCACCGGCGAGGCCACATCAACGTGCCGTAT  1566

Qy   199  ---GlnAlaAlaProSerValPheGlnGlnAlaLeuGlnAlaLeuGlnArgAlaMetLeuGly  217
                    :::        |||:::|||:::||||||:::
Db  1567  CATCAGCCGAAACCAAGGACACTG---AAGGAGTTTCTTAGCCGACGAACCATC---  1617

Qy   218  LysAsnAlaPro---------ValAlaGlyValGlyArgGlyValGluGlnIleArgMetMet  233
               ||||||                :::||||||
Db  1618  ---AACCCTCCACTGGCCACCGGCTCTGGCTGGAGAGTCCAATGCCAGGCAGCAGCCA  1674

Qy   234  MetAsnArgValAspMetGlnArgMetGlnLeuArgGlnLeuGlnGluGlnLeuGluAspAsp  253
               |||||||||        :::   |||
Db  1675  AGGAAGTCGGTGGCAGTGCGGATGACCAGGGAGAATTGGAGGCTTATGCAAAACTTATG  1734

Qy   254  AspAspLeuGluAspGluAsnGluValProAlaArgArgSer--------SerAspGlyGlu  270
               :::||||||                          |||:::
Db  1735  GAGGATCGTGCCAAAGAAGCTACCGAGTTCTTAAGTCTGAGTCTGAGTCAGGATGAGGAG  1794

Qy   271  ProGlnSerGluAlaGluHisGlnArgArgAspLeuArgArgLeuLysSerSerPro  290
               :::||||||
Db  1795  GATGATTCGGAAGAACGAAGAA-----------CCTATGGAATTAAAAGACAATCCA  1839

Qy   291  ArgLeuLysGluLeuLeuGlnAsnAlaGluValGlnSerLeuLeuSerTyrGlnArgMet  310
                                                            GAGATG
Db  1840  GGTGTTATGGAT-----------------------------------GAGATG  1857

Qy   311  ArgAspSerProLeuSerLysArgArgProLeuAlaMetAsnAspGluGluSerAla  330
               ||||||:::   |||
Db  1858  CTGGATAATCCAAAAACTCCAGAACAGCCC----------AAGAACGACACAGAG-  1902

Qy   331  PheArgAlaMetGluAlaMetGluAlaLysAlaLysLeuAspGlnLysSerGlnLeuValLeuGlyLyrLeu  350
                                                   :::
Db  1903  -----------------------------------GACATGACCAGCCAAGCAGTTATTGAGGTA  1932

Qy   351  HisGlyPheGlyGluSerAspAspAspGluGluGluGluAspGluAspGluAsnLeuIle------  368
               |||
Db  1933  CCTGTCCTGGGTGAAGACAAAAACAGATGATACCGTCGGTGATGAAGCAATGGTTGCGAT  1992

Qy   369  ---------AspProSerGluAsnSerPheArgAlaProLeuArgLeuSerGlur  383
               :::|||
Db  1993  TCTATTACTGAAGAAGAACCCATTGCCTACCTCTACTGACCAGTCTGCCTGAAATTAGCA  2052

Qy   384  SerGlyPheValGluLysLysSerAsnAspArgLysLeuLysSerAlaLeuAspAspArgIle  403
               |||:::        |||
Db  2053  GATAATTTGAAATTCAGCAAGAGACTAATGCGAAATCCCCAGCAAAGTGTCTCTGTT  2112

Qy   404  LysTyrArgValAspAspValGluLysIleTyrLeuAlaProLys---------ProMet  419
               |||:::                                          :::
```

XX
DR   WPI; 2001-655860/75.
DR   P-PSDB; ABB59801.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signaling and cell-cell
PT   interactions.
XX
PS   Claim 1; SEQ ID NO 6194; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins (ABB5777-
CC   ABB72072). The sequence data for this patent did not form part of the
CC   printed specification, but was obtained in electronic format directly
CC   from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 5756 BP; 1776 A; 1378 C; 1460 G; 1142 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:   Length: 5756
Score: 168.00   Matches: 157
Percent Similarity: 33.76%   Conservative: 108
Best Local Similarity: 20.00%   Mismatches: 269
Query Match: 5.10%   Indels: 253
DB: 4   Gaps: 35

US-10-736-868-2 (1-643) x ABL03904 (1-5756)

```
Db  2113 ACAGAAGTAGTGGAA---------------------TTGCCCAAGCTGGACTTGAGCACCATC 2154

Qy    39 LysSerHisGlnLysLeuProLeu------------------------------AlaProSer 49
                  :::|||||  ||||||                              |||
Db  2031 AAAGACAACAAGAAGATCACTGATTGCGAGAGATTCCAAACCAAAGAAGTCCTAAG 2090

Qy   420 GluPheAsnProLysProGlnProGlyTyrPheAlaProArgLysIleProThrArgPro 439
                  ::|||||                                           |||
Db  2155 AACATAAACACCGTCCTCTAAGCCA--------------------------------- 2178

Qy    50 MetGluAlaLeuGluLeuMetGlyValGlnPheValAspAlaLeuIleLeuLysLysGlyGln 69
                  |||||||                                      |||||||
Db  2091 AAGGAAACAAGCAAACCGAGGA------------------------AAAGGCAAG 2123

Qy   440 ArgLysMetLeuProLeuLeuIleGlySerAspProLysValGlnGluIleLeuArgArg 459
                  :::  ||||||||||:::  |||  ||||||||
Db  2179 -------GCCAACTTCCTAAGATTAGCGAAGTTATTCGC-- 2208

Qy    70 MetGluMetAlaLysValGlyAlaPheLysThrGlnLeuGluValLeuGluLysValHisPro 89
                  ||||||||                                 :::::|||
Db  2124 AAGGAGGAGAGCAAAACCAGCTGAAAAGAGTAAGAAAGAAGAAGAAGGAA 2171

Qy   460 HisProSerThrGluTrpLysIleAlaLysGluSerArgValLeuThrAsnLeuLysAsn 479
                  |||||||:::||||||
Db  2209 ---------CAGCTGAAGATTGAGAAGAGCTTAGATGAGAGCCCTTCGCTGAAAGGC 2256

Qy    90 AspGlnPheAspLysProLysTyrLysLysLeuLysValLysAspAspLeuAlaAlaAspAlaAlaValAlaMet 109
                  |||  |||                                       |||  :::
Db  2172 GACAGCACACAACAAGTCAAAGAAAGAAGAAGCAGAGACTCACCTGCAAATAATCAAAAGAA 2231

Qy   480 AsnProSerLeuAlaAla------------------LeuPhe---------------- 487
                  :::||:::::::                    |||
Db  2257 GATCCCAACCATGGTTATTGATCTGGAAACGGGTGATATGTTCGCCAAAAACCAACAGT 2316

Qy   110 GlnGln------------------------------------------ 111
                  |||
Db  2232 CAACAAATTGGAGTCAGAGAAAAGACTTCTGAACCAGAAGATGCATCACACAAA 2291

Qy   488 MetAspAspLeuGluAsnThrLysLysGlyArgGlnMetLeuThrAspGluGlnLys 507
                  :::|||||||  |||  |||  |||  |||
Db  2317 GTGGATGAC-----------CTGCTAAACCGTCTGATGAAAACGGCGGAAGCTAAA 2361

Qy   112 AlaGluMetAlaLysLeuLysGlnProLysSerGlyAsnAlaPheIleAspMetLeuAsnGly 131
                  :::  |||||
Db  2292 TCGTGACGGCCAAAAAAGGATTCCAAGAAAAGACGAGGCATCTCACTCACGTTAAGTCGGAC 2351

Qy   508 GlyArgThrArgValLysThrIleArgAlaLeuProArgLeuPheGlyAlaProThrAla 527
                  :::||||||
Db  2362 AAGCACACAAGACCACAGAGACTGTCAACATCCTAACCACTGAACATGGC------------ 2409

Qy   132 AsnGlyIleProIleGlySerSerIleArgGlyLeu------------ 143
                  |||  :::  |||
Db  2352 AACGAAATCTCCGGAAGCAGCAGAAATCTGATGATCCTCCGACCAGCGAGGCCAACAGT 2411

Qy   528 LysAlaGluMetIleAspAlaLysValPhe---------GlnAspIleGluArgPro 544
                  ||:::|||||||                ::::::|||  :::  |||
Db  2410 AAGCTGGAGATGTCCAAGTGACATACACCTGCACGAAGAGAAATCGCAAAGACCA 2469

Qy   144 -------GluAspAlaIleArgThrArgAspMetGluAsnThrAsp---ProSer 159
                  |||  :::
Db  2412 TCCAGCAGCGAGCACGAGATGGATGCGGATACGGACACCGGCACGGACCCTCC 2471

Qy   545 IleProProLeuPheGluProLysArgHisThrArgLeuArgTrpThrGlyThrAla 564
                  |||  :::  :::
Db  2470 AAACCAGGACGCGGTACGGGTTACATGAAAATGCAGGAGCATCCAAA----------ACCCTGATT 2520

Qy   160 GluGlnIleIleAlaLysAlaAlaValMetAspLysPheGlnIleIleLeuProGlyLeuVal 179
                  |||  :::  ||||||
Db  2472 GCCCAA-----------AAGGAGACTCGCCTGAACACAGGAAGAGTCTGCCGGAA---GTA 2516

Qy   565 AsnGluLysGluIleProGlyLeuProGlySerArgPheIleLeuProSerLeuAspProThr 584
                  ||||||  :::  :::
Db  2521 ACTAAAAAAACGCATGGAGGACCTACGCAAAAG------------------------- 2553

Qy   180 AlaAsnMetIleAlaGlyLysAsnPro------PheLysMetProGlnMetArgLys 197
                  |||  :::|||                   :::::  :::|||
Db  2517 ACCGGCGCCCCCGAAGGGTCTTACGCCCAAACACAGCAACGACTCATGGAGCAAGAAAAAG 2576

Qy   585 MetProAlaLeuAsnThrAlaPheSerThrGlnGlyArgAlaArgArgAspGlyTrpAspThr 604
                  |||  :::  |||||||
Db  2554 ----------CAGGCTGAAGAACAGGAGAAAATGGCAGAAGATGAGGAGGAGGC 2598

Qy   605 MetPheLysIleProAsnAsnTrpAsnProGlyAspGluValGlyPhe 620
                  |||  :::  :::||||||
Db  2599 ATG--GATGTGGATGAGGAGTACGAACCGGAAGATAAACCTGGCTAT 2643
```

RESULT 14
ABL03904
ID   ABL03904 standard; cDNA; 5756 BP.
XX
AC   ABL03904;
XX
DT   26-MAR-2002 (first entry)
XX
DE   Drosophila melanogaster expressed polynucleotide SEQ ID NO 6194
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical; gene; ss.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US009231.
XX
PR   23-MAR-2000; 2000US-0191637P.
PR   11-JUL-2000; 2000US-0614150.
XX
PA   (PEKE) PE CORP NY.
XX
PI   Venter JC, Adams M, Li PWD, Myers EW;

```
Qy   198 AlaGlnAlaAlaProSerValPheGlnAlaGlnAlaLeuAlaGlnAlaMetLeuGly 217
             |||     ||||||
Db  2577 GCTCGGCAGGAGAAG------------GAACAAAGCTGGGCGGAGGAGGCGGTCGCCTCAAG 2624

Qy   218 LysAsnAlaProValAlaGlyLeuArgGlyGluGlnLeuArgMetMetMetAsnArgVal 237
             :::                              ||||:||:::
Db  2625 CAACAG----------------GATAAGGAGCACACGGAGCACAAAGAAACAGGAGCGC--- 2669

Qy   238 AspGlnArgMetGlnGlnGluLeuGln---GluGluAspGluAspAspAspLeu 256
             :::::   ||||||||  ::: ||| |||
Db  2670 GACGAAAAGGAACAGCAGCGGAAGCTCGAGAAGGGATCAGAAGGCAGCAGAGGAAGATG 2729

Qy   257 GluAspGluGluAspGlyValProArgArgArgSerSerAspGlyGlnSerGluAlaGlu 276
             |||:   |||||::    ::::  :::|||
Db  2730 GAAAAGGAGAGGAAGAAGAGAGGCCAAGCAGGCCGAGCTGGACTCCAAGAACGAGGAGAAG 2789

Qy   277 HisGlnArg----------ArgAspLeuAlaArgArgLeuLysSerSerProArgLeuLys 293
             :::||     :::::   |||   |||
Db  2790 CGCAACGGCCAACGAGGAGGAGGTGCAACGCAAAAGGAGCACGGAGGCGTCGCAAG 2849

Qy   294 GluLeuLeuGlnAsnAlaGluValGln--- 302
             :::  :::  |||
Db  2850 AAGGAGCAGGACGCGTGAGGAGGCCCGACCAAAGAAGAAGCGCCCGAGTCCTTCTCC 2909

Qy   303 SerLeuLeu 305
             |||
Db  2910 AAGTTCTTCGTGCCCAAGCAGCCCAGTGCGGCGGATCCAACAACATCCTACCTG 2969

Qy   306 SerTyrGlnArgMetArg--AspSerProLeuSerLysArgArgProLeuAlaMetAsn 324
             :::::::::   ||||||   :::   ::: |||
Db  2970 GAGCACGCAAAGCAGTCGCCAGCTGCCCAAGGGCCAGCGACCAGCCCAAACGCTG--- 3020

Qy   325 AspGluAspGluSerAlaPheArgArgAlaMetGluAlaArgAlaAlaLysLeu--- 340
             ||||||   |||
Db  3021 --------GCCTTTCGACCATTTCAGATCAAGGACGACATGCTGCTAGCACCG 3065

Qy   341 AspGlnLysSerGlnLeu--ValLeuLeuHis 351
             :::||::|:|::||||||||
Db  3066 ATCGTCGCAACTCGCTCGGCAGGAGCAGCGGCGCCCCAGCTAGATGACTTTTTCGACAC 3125

Qy   352 GlyPheGlyGluSerAspAspAspAspGluAspGluAsnLeuIle-------- 368
             ::::|||||||:::|||:|||||||||:::
Db  3126 CGGGACGAGCGAAGCTGCGACGAGGAAGAGGCAGACATCGTCGACGACGAAAG 3185

Qy   369 AspProSerGlu-----------AsnSerPheArgArgAlaProLeu 380
             |||||:::   ||||||:::||||||
Db  3186 CCGGCCGAATCGGGCTAATCTCTATCTTAGTGAGTTGAGCAGCCGCAAGCCGCTC 3245

Qy   381 ArgLeuSerSerGlyPheValGluLysLeuLysLeuLysSerAsnAspGlu- 395
             :::::::   |||   :::   ||||||||
Db  3246 AAGATGCAGCGCGATGTGAAGCTCCAAAGACCACCAGGACGAGGATGATGACGAT 3305

Qy   396 -----------------LeuLysSerAlaLeuAspArgArgIleLeuLys 404
             :|::|||     |||
Db  3306 GTTCAAGTCGGATAGGTAGGATATGGCTGCAAGATATTAAAGAAGCATTT-------GTTCTT 3359

Qy   405 TyrArgValAspAspValGluLysLysTyrLeuAla--------- 416
             :::::: |||||||||   ||||||
Db  3360 CACCAATCATCTGCATCTACAGATTACCTTTCGCCGGGTCTCCCATCGAGGTGGAG 3419

Qy   417 --------LysProMetGluPheAsnProLysProGln 426
             :::          |||
Db  3420 CAGCCCAAGCAGCCGGAATGGAGCCGGAGTATCTGCATTTCGGCCGACAACCCGGAGG 3479

Qy   427 ProGlyTyrPheAlaProArgLysIleProThrArgProArgLysMetLeuProLeu 446
             :::::::          :::
Db  3480 CCACCGTACTACGGCCACCTGG----------------CGTAAAAGAGCAGCAGCATC 3521

Qy   447 IleGlySerAspProLysProLysValGlnGluGluIleLeuArgArgHisProSerThrGluTrpLys 466
             ::::::                                ||||||
Db  3522 TCCGCCCCAGCGCCGCCTGGCACAAGGATAAGGTA--------------TGGATC 3560

Qy   467 IleAlaLysGluSerArgValLeuThrAsnLeuLysAsnProSerLeuAlaAlaLeu 486
```

```
Qy         :::                   ||| :::::|||
Db  3561 ACTTCTTCTGGCTTAAATATCATAACACAACAAATATC-AATAATCATATAACGCAGTGCTC 3619

Qy   487 Phe-----------------------------MetAspAspLysLeuGluAsnThrLeuLysGly- 498
             |||                           |||
Db  3620 TTCGACTACGAAGTGGACTCCGATTGCGATTGGGACTGGGACGAGGAGGGGGAATCCGTTG----------- 3679

Qy   499 ------------ArgGlnMetLeuThrArgGluGlnGlyGlnLysGlyArgThrArgVal 512
                                   :::::||||||  :::  ::::
Db  3680 AGCGCTAGCGAGGACGAGGAGAAAGAAGAGAGTCCGAGGAGGAATC-GGAAGAGGAGTACAA 3738

Qy   513 LysThrIleArgAlaGlu- 518
             :::::  :::::||||||
Db  3739 CGAGTGGTACGTGCCGCACGGACATCTCAGCGACGAGGAACTGCAAAACGACGGCGACGG 3798

Qy   519 -----------------ProThrAlaLysAlaGlu- 530
             |||                  |||
Db  3799 AATGGAAGACGGCCACACGCGGAGGCGCAAACTCCAGTGCTGCAGCAGGA 3858

Qy   531 MetIleAspAlaLysValPheGlnAspIleGluGluPro-------- 546
             :::                        :::::|||||||
Db  3859 GTTCGGCTCCAGGAAATGAAGAAGCAGACGAAAAAGATCAAGGGCGCCGCCTGTTGGGTCCCGT 3918

Qy   547 ----------------ProLeuPhePheGluProLysGly 554
                              |||||||
Db  3919 TTGGCTGGACGAACGGCAACAAGTCCAGCTATTCCCAGCTATCTTTGCCCACCAT 3978

Qy   555 ArgHisThrArgLeuArgTrpThrGlyAlaAsnGluLysGluIleProGlyLeuGlySer 574
             ::::|||||||||||                              :::
Db  3979 CGACAATGTACGGCTTGGTTGGCAGTGGAGGCCGTCCGGAGCCCACTCAG--------CCTAGAGCCTCCGCC 4029

Qy   575 ArgPheIleLeuProSerLeuAsp- 583
             |||                   |||
Db  4030 AGAACCGGAGGGTCAGGATCAGAGCAGCCCGGTGCCGCTCGAGCTGCAGCGATCGG 4089

Qy   584 ThrMetProAlaLeuAsnThrAlaPheSerThrGlnGlyAspArgAlaArgAspAspGluTrpAsp 603
             |||                              ::::::
Db  4090 TCTCATGCAGCAGCTGGTGCGGCCTTACCCACGGCAATCGGAACTCGAAGACGTTCCTGAT 4149

Qy   604 ---ThrMetPheLysIleProAsnAsnTrpAsnProGlyAspGluValGlyPheLysMet 622
             :::                                        :::
Db  4150 CAACGGAATATTTAGAGTACCTGAA-------AACCCAAGCAGCGACCGCCCAGGCCAATCAAAC 4203

Qy   623 AsnSerLysThrLys 627
             |||
Db  4204 AACGCTGCCCTCAAA 4218
```

PI   Venter JC, Adams M, Li PWD, Myers EW;
XX
DR   WPI; 2001-656860/75.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signaling and cell-cell
PT   interactions.
XX
PS   Claim 1; SEQ ID NO 37324; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL01840-ABL16175), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
CC   ABB72072). The sequence data in this patent did not form part of the
CC   printed specification, but was obtained in electronic format directly
CC   from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 6968 BP; 1858 A; 1845 C; 1933 G; 1332 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:   Length: 6968
Score: 0.003   Matches: 136
Percent Similarity: 167.50   Conservative: 98
Best Local Similarity: 36.17%   Mismatches: 204
Query Match: 21.02%   Indels: 209
DB: 5.09%   Gaps: 29

US-10-736-868-2 (1-643) x ABL28617 (1-6968)

Qy   24 SerThrSerAspAlaProGlyAlaLeuLeuSerSerLeuValGlyLysSerHisGlnLys 43
             ||| |||||| ::: ::::: ||||| :::::|
Db   4531 AGCAAGTACGAGTCCGATGGCGTTGCCCGCTCTGAGGAGCTGGAGGAGCCAAGAGGAAG 4590

Qy   44 LeuPro------LeuAlaProSerMetGluAlaLeuGluLeuMetGlyValGlnPheVal 61
             |||| ||| ||||| |||||||||:
Db   4591 CTGCAGGCCGCGTTTGGCCGAGGGCCGAGGAGACCATGAGTCCCTCAACCAGAAGTGCATT 4650

Qy   62 AspAlaLeuIleLysLysGlyGlnMetGluMetAlaLysAlaLysGlyAlaPheLysThrGlnAsn 81
             :::||| ||||| |:::::::|||||||||
Db   4651 GGC----------CTGGAGAAGACCAAGCAGCGCTGTGTCCACCCAGGTG 4689

Qy   82 GluValLeuGlu---------------------LysValHis 88
             ||| |||:::   :::
Db   4690 GAGCTGGAGCTGTCCGAGGTCGAAGGAGTCGCAACTACTCCACCAGCTGTTCCGTCTTAAG 4749

Qy   89 ProAspGlnPheAspLysTyr----------LysLysLysLysValAspAspLeuAlaAla 105
             ||::|||||||| |||:::
Db   4750 CAGAAGGCCTTCGACAAGATCATCGGCGAGTCAAGGTCGACGATCTGGCTGCT 4809

Qy   106 AspAlaValMetGlnGln------------AlaGluMetAlaLysLeuGln 118
             ::: ||||| :::::
Db   4810 GAGCTGGATGCCTCCCCAGAAGGAGTCGCCGCAACTACTCCACCAGCTGTTCCGTCTTAAG 4869

Qy   119 ProLysSerGlyAsnAlaPheIleAspMetLeuAsnGlyAsnGlyIleProIleGlySer 138
             |||| :::|||
Db   4870 --GGCGCC 4875

Qy   139 SerIleArgGlyLeuGluAspAlaIleArgThrGlnAsnArgMetGluAsnThrAspPro 158
             |||||||:::
Db   4876 TACGGAGGAGGGCCCAGGAGCAGTTGGAGGCTGTCGTCGTGAGGACAAGAAC-------- 4926

Qy   159 SerGluGlnIleAlaLysAlaValMetAspLysAspLysPheGlnIleLeuLeuProGlyLeu 178
             ::||| |||
Db   4927 ----CTGCCGATGAGGTCAAGGATCTGTCCGAC---CAGATCGGTGAGGGT---- 4971

Qy   179 ValAlaAsnMetIleAlaGlyLysAsnProPheLysMetProGlnMetArgLysAla 198
             |::||||| ::::: ||||| |||:::
Db   4972 ------------------------------GGCCCAACATCCATGAGATC---GAGAGGAGCCAAGCGC 5010

Qy   199 GlnAlaAlaProSerArgValPheGlnGlnAlaLeuAlaGlnArgAlaMetLeuGlyLys 218
             ||| |||||| |||||| :::
Db   5011 CTGGAGAGCCGAGAAGGACGAGGACCTGCCAGGCTGCCCCTGCCCTGAGGAGGCTGAG---- 5058

Qy   219 AsnAlaProValAlaGlyGlyArgGlyArgGlyValGlyGluGlnArgMet----------- 233
             |||||| :::::::::
Db   5059 --------GCCGCCTTGAGCAGCAGCGAGGAGAACAAGTGCTCCGGCTTCAGCTTGAG 5106

Qy   234 MetAsnArgValAspGlnArgMetGlnGlnLysArgGluLeuGlnGluLysGluGluAspAsp 253
             :::::::||| ::: |||:::|||||| :::||||||||:::
Db   5107 CTGTCCCAGTGCGCCCAGGAGATCGACCCCGC--ATCCAGGAGGAGGAGGAGGAGTTC 5163

Qy   254 AspAspLeuGluAspGluValProArgArgSerArgAspGlyGluProGlnSer 273
             ||||||| |::::::::
Db   5164 GAGAACACCCGCAAGAACCACCACGGTGCCCTGACTCCATGCAGCTTCCCTGGAAGCC 5223

Qy   274 GluAlaGluHisGlnArgArgAspLeuArgArgLeuLysSerSerProArgLeuLys 293
             ::|||||||| ::: |||:::
Db   5224 GAGGCCAAGGGCAAGGCTGAGGGCCTGCGCATGAAGAAGAGCTGGAGGCTGACATCAAC 5283

Qy   294 GluLeu---------------------------LeuGlnAsnAlaGluValGln------ 302
             ||||| |||||||
Db   5284 GAGCTTGAGATTGCTCTGGATCACGCCAACAAGGCTAACGCCCAGGAGCCAGAAGAACATC 5343

Qy   303 ----------------------------------SerLeuLeuSerTyrGlnArgMet 310
             :::||||
Db   5344 AAGGGTTACCAGCAGCAGTCTGAAGGAGACATCCAGCCTGAGGAGGAGCAGCGCGCC 5403

Qy   311 ------------LeuSerLysArgArgProLeuAlaMetAsnAsp 325
             |||:::||||||
Db   5404 CGCGGACGATGCCCCGGAACAGCTGGGTATCTCCGAGCGTGCGTCAAACGCCCTCCAGAAC 5463

Qy   326 GluAspGluSerAlaPheArgAlaMetGlu------------- 335
             :::|||||| :::
Db   5464 GAACTGGAGGAGTCTGCACTCTGTGGAACAGGCCGACCTGGCCGTCGGCCGGAA 5523

Qy   335 --- 335

Db   5524 CAGGAGCTGCCGATGCCCAGCAGCTGAACGAGCTGAACAAGAGTGTCCGCCCAGAGACGCCTCCATC 5583

Qy   336 ------AlaArgAlaLysLeuLysAspGlnLysGlnLysSerGlnAsnLysSerGlnGlyLeuHisGlyPhe 353
             ||| ||| ::::::::::
Db   5584 TCCCTGCCAAGAGAGGAAGCTGGAGTCCGAGCTGCAGACC-----------CTGCAC-- 5628

Qy   354 GlyGluSerAspAspAspValGluAspGlyAsnLeuIleLeuAspProSerGluAsn 373
             ||||||||| :::::::::
Db   5629 -----------TCCGACCTGGACGAA----CTCCTGAACGAAGCCAAGAAC 5664

Qy   374 SerPheArgArgAlaProLeuArgLeuSerSerGlyPheValGluValLeuLysSerAsn 393
             ||| ::::::: ||| :::
Db   5665 ----------AAGAAGGCTATGTCATGGTGCCGTCTGCGACGAGGCTGGCC 5712

Qy   394 AspGluLeuLysSerAlaLeuAsp------------------------------------ 401
             ||||||||||| |||
Db   5713 GATCAGCTCCGCGTCGAGGACGATCATGCCCAGGAGAAATTGACGAGGCCCTC 5772

Qy   402 ------------------------------ArgIleLysTyrArgValAspAspValLysTyrLeuAlaPro 416
             ::::::: |||:::||||||
Db   5773 GAGCAGCAGATCAAGGAGCTGCAGGTCCGTCTGACGAGGCTGGCCAACGCCCCTC---- 5829

Qy   417 LysProMetGluPheAsnProLeuProGlnProGlyTyrPheAlaProArgLysIleProPro 436
             ::: ||||||
Db   5830 --------AAGGGAGGCAAGAAGGCCATTCAGAAGCTTGAG 5862

Qy   437 ThrArgProArgLysMetLeuProLeuLeuIleGlySerAspProLysValGlnGluGlu 456
             |||::: |||:::
Db   5863 CAGCGCGTCCGCCGAGCTC---------GAGAACGAGCTGGATGGTGAG 5901

Qy   457 IleArgArgHisProSerPheGluTrpLysIleLeuAlaLysGluSerAlaValLeuThrAsn 476
             |||||| |||:::
Db   5902 CAGAGGAGGCACCCGCGATGCCCAGAAGAACCTGCGCAAGTCCAAGTCCAAGGAG 5961

Qy   477 Leu--------------------------LysAsnAsnProSerLeuAlaAlaAlaLeuPheMet 488
             ||||| |||:::
Db   4972 --------------------------GGCCCAACATCCATGAGATC---GAGAGGAGCCAAGCGC 5010

```
Db      |||
     5962 CTGAGCTTCCAGTCCGAGGAGGACCCGCAAGAACCACGAGCGCATGCAGGATCTGGTC--- 6018

Qy          |||||||:::                        :::                    |||
      489 AspAspLysLeuGluAsnThrLeuLysGlyArgGlnMetLeuThrAspGluGlnLysGly 508
            ||||||||||:::::                    :::|||
Db   6019 ---GACAAGCTGCAACAGAGATCAAGACATACAAGAGGCAGATCGAGGAG-------- 6066

Qy      509 ArgThrArgValLysThrIleArgAlaLeuProArgLeuPheGlyAlaProThrAlaLys 528
            :::  |||  |||||||    |||           |||:::
Db   6067 --------GCTGAGGAAATCGCCGCCCTC---AACTTGGCCAAATTCCGCAAGGCTCAG 6114

Qy      529 AlaGluMetIleAspAlaLysValPheGlnAspIleGluGlyArgProIleProProLeu 548
            ||:::  :::|||::      |||::::       ||:::     |||
Db   6115 CAGGAGCTTGAGGAGGCCGAGAGCGCCGCCGATCTCGCCGAGCAGGCCATCAGCAAA--- 6171

Qy      549 PhePheGluProLysGlyArg 555
            |||     |||||||||
Db   6172 ---TTCCGCGCCAAGGGACGT 6189
```

Search completed: July 15, 2005, 16:01:48
Job time : 1204 secs

```
          GenCore version 5.1.6
    Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on:      July 18, 2005, 12:25:31 ; Search time 305 Seconds
                                        (without alignments)
                                        3449.593 Million cell updates/sec

Title:           US-10-736-868-2
Perfect score:   3291
Sequence:        1 MILFLFLLLGFCIAPLSA.........SKTKRFVGGNGAFDMPALGL 643

Scoring table:   BLOSUM62
                 Xgapop 10.0 , Xgapext  0.5
                 Ygapop 10.0 , Ygapext  0.5
                 Fgapop  6.0 , Fgapext  7.0
                 Delop   6.0 , Delext   7.0

Searched:        1202784 seqs, 818138359 residues

Total number of hits satisfying chosen parameters:        2405568

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=x1h
-Q=/cgn2_1/USPTO_spool/US10736868/runat_14072005_105510_2303/app_query.fasta_1.839
-DB=Issued Patents NA .-QFMT=fastap -SUFFIX=rml -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10736868 @CGN_1_1_69 @runat_14072005_105510_2303 -NCPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=10 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :   Issued Patents NA:*
         1:  /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
         2:  /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
         3:  /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
         4:  /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
         5:  /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
         6:  /cgn2_6/ptodata/1/ina/backfiles1.seq:*
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	171	5.2	100990	4	US-09-409-800B-2	Sequence 2, Appli
2	169.5	5.2	6151	4	US-09-799-451-528	Sequence 528, App
3	162.5	4.9	5883	4	US-09-949-016-5001	Sequence 5001, Ap
4	159	4.8	2680	2	US-08-533-306A-5	Sequence 5, Appli
5	159	4.8	2680	2	US-08-742-923A-5	Sequence 5, Appli
6	159	4.8	2887	2	US-08-533-306A-3	Sequence 3, Appli
7	159	4.8	2887	2	US-08-742-923A-3	Sequence 3, Appli
8	159	4.8	4935	4	US-09-949-016-1019	Sequence 1019, Ap
9	159	4.8	4935	4	US-09-949-016-5054	Sequence 5054, Ap
10	159	4.8	5574	4	US-09-917-254-40	Sequence 40, Appl
11	159	4.8	6861	4	US-09-949-016-1240	Sequence 1240, Ap
12	159	4.8	6861	4	US-09-949-016-1241	Sequence 1241, Ap
13	159	4.8	6861	4	US-09-949-016-1242	Sequence 1242, Ap
14	158.5	4.8	19269	4	US-09-902-540-1175	Sequence 1175, Appli
15	157	4.8	6175	4	US-08-875-435B-1	Sequence 1, Appli
16	157	4.8	6856	4	US-09-566-921-42	Sequence 42, Appli
17	156	4.7	3807	4	US-09-645-456A-8	Sequence 8, Appli
18	156	4.7	3807	4	US-09-645-324A-8	Sequence 8, Appli
19	156	4.7	3807	4	US-09-645-791-8	Sequence 8, Appli
20	155.5	4.7	3972	4	US-09-645-456A-6	Sequence 6, Appli
21	155.5	4.7	3972	4	US-09-425-324A-6	Sequence 6, Appli
22	155.5	4.7	3972	4	US-09-645-791-6	Sequence 6, Appli
23	153.5	4.7	6396	4	US-09-949-016-3344	Sequence 3344, Ap
24	153	4.6	9551	1	US-08-056-200-93	Sequence 93, Appl
25	153	4.6	9551	2	US-08-800-644-93	Sequence 93, Appl
26	152.5	4.6	1761	4	US-09-252-991A-16401	Sequence 16401, A
27	152.5	4.6	3489	4	US-09-252-991A-16193	Sequence 16193, A
28	152.5	4.6	4055	4	US-09-688-188B-10	Sequence 10, Appl
29	152.5	4.6	4055	4	US-09-291-417D-10	Sequence 10, Appl
30	152	4.6	3831	4	US-09-645-456A-5	Sequence 5, Appli
31	152	4.6	3831	4	US-09-425-324A-5	Sequence 5, Appli
32	152	4.6	3831	4	US-09-645-791-5	Sequence 5, Appli
33	151.5	4.6	3996	4	US-09-645-456A-2	Sequence 2, Appli
34	151.5	4.6	3996	4	US-09-425-324A-2	Sequence 2, Appli
35	151.5	4.6	3996	4	US-09-645-791-2	Sequence 2, Appli
36	151.5	4.6	6644	4	US-08-875-435B-5	Sequence 5, Appli
37	150.5	4.6	3228	4	US-09-252-991A-1816	Sequence 1816, Ap
38	150.5	4.6	3546	4	US-09-252-991A-2143	Sequence 2143, Ap
39	150.5	4.6	4209	4	US-09-248-796A-400	Sequence 400, App
40	150.5	4.6	7596	4	US-09-023-655-1463	Sequence 1463, Ap
41	149	4.5	2130	4	US-09-248-796A-4942	Sequence 4942, Ap
42	148.5	4.5	2094	4	US-09-248-796A-1868	Sequence 1868, Ap
43	146.5	4.5	3894	4	US-09-645-456A-7	Sequence 7, Appli
44	146.5	4.5	3894	4	US-09-425-324A-7	Sequence 7, Appli
45	146.5	4.5	3894	4	US-09-645-791-7	Sequence 7, Appli

ALIGNMENTS

```
RESULT 1
US-09-409-800B-2

; Sequence 2, Application US/09409800B
; Patent No. 6706522
; GENERAL INFORMATION:
;  APPLICANT: Blattner, Frederick R.
;  APPLICANT: Burland, Valerie
;  APPLICANT: Rose, Debra J.
;  APPLICANT: Mayhew, George F.
;  APPLICANT: Perna, Nicole
;  APPLICANT: Perry, Robert D.
;  APPLICANT: Straley, Susan C.
;  APPLICANT: Fetherston, Jacqueline D.
;  APPLICANT: Lindler, Luther E.
;  APPLICANT: Plano, Gregory V.
;  TITLE OF INVENTION: Plasmid DNA From Yersinia Pestis
;  FILE REFERENCE: 960296.95939
;  CURRENT APPLICATION NUMBER: US/09/409,800B
;  CURRENT FILING DATE: 1999-09-30
;  NUMBER OF SEQ ID NOS: 3
;  SOFTWARE: PatentIn version 3.1
;  SEQ ID NO 2
;  LENGTH: 100990
;  TYPE: DNA
;  ORGANISM: Yersinia pestis
US-09-409-800B-2

Alignment Scores:
Pred. No.:              2.82e-05        Length:        100990
Score:                  171.00          Matches:       142
Percent Similarity:     33.99%          Conservative:  100
Best Local Similarity:  19.94%          Mismatches:    260
Query Match:            5.20%           Indels:        210
DB:                     4               Gaps:          33

US-10-736-868-2 (1-643)  x  US-09-409-800B-2 (1-100990)
```

```
Qy     17 ProLeuSerAlaGlnSerProSer---ThrSerAspAlaProGlyAlaLeuSerSer        35
             |||||| :::||||
Db  85225 CCAGTGCTGGTAAACATCCCATCTATACCGGACGACGCATCACCACCCTCATGAACGCA  85284

Qy     36 LeuValGly-                                                      38
             |||:::|||||
Db  85285 GTACGTGGTTTCTCGACCATGAAGTAGCCCATATCCTCTTTACTGATCCGAAAGTCGCG  85344

Qy     39 ----LysSerHisGlnLysLeuProLeuAlaProSerMetGluAlaLeuGluLeu        55
               ||||||
Db  85345 ATGAAGAGATGCCGAGAGAAGAGGCTCCGTCTACCGGGCTTTGGAACGGCGCTGGAA  85401

Qy     56 MetGlyValGlnPheValAspAlaLeuIleLeuLysGlyGlnMetGluMetAlaLysGly     75
             |||
Db  85402 ---------GACGTTTTATTGAACGGACGAGAATGGGACAGGTATTCAACGGA  85443

Qy     76 AlaPheLysThrGlnLeuGluLeuValLeuLysValHisProAspGlnPheAspLysTyr     95
             :::
Db  85444 ACCCGTCGTAATTTGCTAGCCACACAGAACCTGGTG-------ATCGACAAATAC  85491

Qy     96 LysLysLeuLysValAspAspLeuAlaLeuAlaValMetGlnGlnAlaGluMet-       114
             ::: :::
Db  85492 TTCAAGGGCAAAGTGTTCAATCTGCCACGGCACCAGCGTGAATTGTTC  85551

Qy    115 -------AlaLysLeuGlnProLysSerGlyAsnAla---PheIleAsp            127
                :::|||||||
Db  85552 CTGAAATTCTTCCTTTGCCCGCGTCTGGGATGGCCAAAGCCCTTTCATCGAC  85611

Qy    128 MetLeuAsnGlyAsnGlyIleProIleGlyTyrGlySerIle----ArgGly        142
             |||
Db  85612 TTTATGGAAGAGACACTGGCACCTCATCGATAAGCCTTGTAGCCTTGCTCAAAGAGCATGGT  85671

Qy    143 LeuGluAspAlaIleArgThrGlnAlaAspMetGluAspProSerGluGlnIle         162
             ::: |||||||
Db  85672 ATCGACGTGGCCGTT-----CGCAATAGTCGAACACAGAGACTGCGTAAGGTC  85722

Qy    163 AlaLysAlaValMetAspLysPheGlnThrGlnIleLeuProGlyLeuValAlaAsnMet    182
             ::: |||
Db  85723 GCAGCTGCTATCGCCCCAAATCATGCAGGAT-------GAGGAAAACACCAGAG  85752

Qy    183 IleAlaGlyLysLysAsnProPheLysMetProGlnMetArgLysAlaGlnAlaAlaPro    202
             ::: :::
Db  85753 ATGAAAGACAAGCCAGAGGGCAAATTACCT-------GAGCTTAAATCATCTGCCAAAAAGCCG  85809

Qy    203 SerSerValPheGlnGlnAlaLeuAlaLeuAlaGlnArgAlaMetLeuGlyLysAsnAlaProVal  222
             :::
Db  85810 TCTAAGAGCAAAGACGAGTCC------GAGGAAAACACCAGAG  85845

Qy    223 AlaGlyLysArgGlyGluGluGlnArgMetMetAsnArgValAspGlnArgMetGln       242
             ||||||
Db  85846 TCAGGGCGACGATCCGGTCTACATGTGAATCTGCACCAAAGGCGAACCAAAGGACGAGACGAC  85905

Qy    243 GlnArgGluLeuGlnGluGluAsp---GluAspAspAspLeuGluAspGluAspValAsp     261
             ||||||
Db  85906 GACAAGGAAGAGACAAGAAGATGATGCCTCAGAAGAGGAAGATTCTGGGGATTCTGATTTG  85965

Qy    262 ProArgArgArgSerSerAspGlyProSerGlnSerGluLeuAlaGlnHisGlnArgArgAsp   281
             ||| :::
Db  85966 GATTCATTAGATAAGGACTTACCTATACATGAT------                        86001
             CCT

Qy    282 LeuAlaArgLeuLysSerSerProLeuLeuGluLeuGlnLeuAsnAlaGluVal         301
             :::|||:::|||
Db  86002 AAGAAATTAGTGATACAGAAAGTAA  86028

Qy    302 GlnSerLeuLeuSerTyrGlnArgMetArgAspSerLysArgArgProLeu           321
             |||:::||||||
Db  86029 TATACAGAAGCAGGCGAAGATGAGTCAGGAGCAACCCAGAATCCGATGATGCCGGCATG  86088

Qy    322 AlaMetAsnAsp--GluAspGluSerAlaPheArgAlaMetGluAla               336
             ||| :::||||||
Db  86089 GAATCAGGTGATTCTGACGACGGAAGGTGGTAGTGATGGCACTGGCCGGCTCCAACGCCTGGT  86148
```

```
Qy    337 ------ArgAlaLysAspLnLysSerGlnLeuValLeuLeuHisGlyPhe             353
             |||
Db  86149 GATGGCAATTCGCGAAGACGCCGATGAACTCCGATGGTTATGGCTCCGGCCGCCTGGCGAT  86208

Qy    354 GlyGluSerAsp                                                   361
             :::||||||
Db  86209 GGGGATAGTGATAACGGCAGACTCCGGTCGGTGTCGGAGGAGAAGAC  86268

Qy    362 GluGluAspGluAsnLeuIleAspProSerGluAsnSerPheArg                 376
             |||
Db  86269 GAAGGCGAAGAGAGGACGTCGCCGATCACACTGACGGCGCAAGGAAAATGAGGATGCC  86328

Qy    377 ---ArgAlaProLeuArgLeuSerSerGlyPheVal- GluLysLeuLysSer         392
               ||||||||||
Db  86329 GCGGAAGCGCCTGAACGACAGTGAGTCAGGCTTTGTTCCTGCCGGATGAAATGACTCTG  86388

Qy    393 AsnAspGluLysLysSerAlaLeuAspArgIleLysTyrArgValAspAspVal------  410
             ||| :::
Db  86389 GAGGACGCACTCAAG--GCGGTCGACGAGTGAAGAGAACGAGAAATGACCGAA  86445

Qy    411 ------------GluLysTyrLeuAlaProLysProMetGluProGluAsn           422
             ::: :::
Db  86446 GACGCACTGTCGGCCACCATCAGCAAAGAGCTTATGAGCCGCCTCACTTTCTGAGTATCGC  86505

Qy    423 ProLysProGlnProGlyTyrPheAlaProArgLysIleProThrArgProArgLysMet   442
             |||
Db  86506 CCATACGATCGTTCATACGACTTT  86529

Qy    443 LeuProLeuLeuIleGlySerAspLysAspProValGlnGluIleArgArg-          459
             |||
Db  86530 ATCGGGTTGATTGATGAGGCTGAAGAGCATGTAAAACGCACCAGAAAG  86577

Qy    460 ----------HisProSerThrGluTrpLysIleAlaAlaLysGluSer             471
             :::
Db  86578 ACATTCGGCAATCCAATGCACTCACCTGTCGATCGCTACCGCATGGTTCCGGAAGGC  86637

Qy    472 ArgValLeuThrAsnLeuLysAsnAsnProSerLeuAla---------Leu          486
             |||:::||
Db  86638 AGAAAAACTCTTTGAACTGAAAAATCGAAAAACATCTGTCTGCCGGGTCGTTCTTCGACTCTG  86697

Qy    487 PheMetAspAspLysLeuGluAsnThrLeuLysGlyArg---------GlnMetLeuThr   503
             ||| |||
Db  86698 GCCAAAGAC-----CTGGAGCGAGCTATCGCCAGCCGCAACCGAGTTCAGTCATACCG  86751

Qy    504 AspGluGlnLysGlyArgThrArgValLysThrIleArgAlaLeuProLeuPheGly     523
             :::|||:::
Db  86752 GGTCAGAGACGTGGG------CGGATACATGGC  86778

Qy    524 AlaProThrAlaAlaGluMetIleAspAlaLeuValPheGlnAsnIleGluGlu--      542
             ||| :::
Db  86779 GCGAACCTGTATCGTCTGGCAATGAACGACCAAAGGAGAAGAGACCAC  86838

Qy    543 --------------ArgProIleProLeuPheGluProGlyLysArg              555
             :::
Db  86839 AGAGCCGTGAACGCGTGCTCGACTTGTCCGGCTCAATGGGCGGCAGG  86898

Qy    556 HisThrArgLeuArgTrpProGlyLeuAlaAsnGluLysProGlyLeuIleProGlySerArg  575
             |||
Db  86899 GCACTCGGCCAGCGCC  86925

Qy    576 Phe----IleLeuProSerLeuAspPro---ThrMetProAlaLeuAsnThrAlaPheSer  593
             :::|||||| ::: |||
Db  86926 GCGAACCTGTATCGTCTGGCAATGCCAAAGGAGAAGATTGCCTATCAATCATCACCGGCTTTACA  86985

Qy    594 ThrGlnGlyArgAlaArgAspGluThrMetPheLys-----------             607
             ::: |||
Db  86986 AGCCCAGATTATGAAACCATGTCGAAGGCGCGGGGTTTACACGTTTC  87039

Qy    608 IleProAsnAsnTrpAsn                                           613
             :|||
Db  87040 GAGGGCGCTCATGCTGCCCATTATCAAAAACTGGAAT  87075
```

RESULT 2

```
US-09-799-451-528
; Sequence 528, Application US/09799451
; Patent No. 6783969
; GENERAL INFORMATION:
; APPLICANT: Tang, Y., Tom
; APPLICANT: Zhou, Ping
; APPLICANT: Goodrich, Ryle
; APPLICANT: Asundi, Vinod
; APPLICANT: Ren, Feiyan
; APPLICANT: Zhang, Jie
; APPLICANT: Xue, Aidong J.
; APPLICANT: Zhao, Qing A.
; APPLICANT: Wang, Jian-Rui
; APPLICANT: Ma, Yunqing
; APPLICANT: Yamazaki, Victoria
; APPLICANT: Chen, Rui-hong
; APPLICANT: Wang, Zhiwei
; APPLICANT: Wang, Dunrui
; APPLICANT: Yang, Yonghong
; APPLICANT: Wehrman, Tom
; APPLICANT: Ghosh, Reena
; APPLICANT: Drmanac, Radoje T.
; TITLE OF INVENTION: No. 6783969el Nucleic Acids and
; TITLE OF INVENTION: Polypeptides
; FILE REFERENCE: 803
; CURRENT APPLICATION NUMBER: US/09/799,451
; CURRENT FILING DATE: 2001-03-05
; NUMBER OF SEQ ID NOS: 948
; SOFTWARE: pt_FL_genes Version 2.0
; SEQ ID NO 528
; LENGTH: 6151
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (263)..(3133)
US-09-799-451-528

Alignment Scores:
Pred. No.:              4.85e-07        Length:        6151
Score:                  169.50          Matches:       118
Percent Similarity:     36.28%          Conservative:  108
Best Local Similarity:  18.94%          Mismatches:    248
Query Match:            5.15%           Indels:        149
DB:                     4               Gaps:          22

US-10-736-868-2 (1-643) x US-09-799-451-528 (1-6151)

Qy   36 LeuValGlyLysSerHisGlnLysLeuProLeuAlaProSerMetGluAlaLeuGluLeu 55
         ||| |||  ||||   |::::|||          ::::::
Db  632 CTGACTGCTCATCCCATCATCACCACCAGGTCCCCTCCATG---TTGAGGCAGGTA 688

Qy   56 MetGlyValGlnPheValAspAlaLeuIleLysLysGlyGlnMetGluMetAlaLysGly 75
         :::||                     ||  ||   |::::::
Db  689 AGAGACAGGCAGGCTAAGGAGCTGTTCCTTTGGAAGAACATTAGAGAGAAATGGAGATGAC 748

Qy   76 AlaPheLysThrGlnLeuGluLeuLysGluLysValHisProAspGlnPheAspLysTyr 95
         :::  |||:::::::           :::
Db  749 CTCCTCCAGCAAGAGAGTGGCTAGACATCAAGGACAGCAAATTGGATCTTCCATGAACAGTATT 808

Qy   96 LysLysLeuLeuLysValAspAspLeuGluLeuAlaAlaAspAlaValMetGlnAlaMetAla 115
         ::::||
Db  809 AAGAACTTTCTGAGTTCTGGAGCTGTAAGGAGGAGGAGAGTCTTGAGGAGGAAGAGGAGCCGCG 868

Qy  116 LysLeuGlnLeuProLysSerGlyAsnAlaPheIleAspPheMetLeuAsnGlyAsnGlyIlePro 135
         :::||
Db  869 CGGATGTCTGTCTCCAAGGAGCAGCAGGATGAGGGGTTTCCCATGAAGAAATCAGCCTTACAG 928

Qy  136 IleGlySerSerIleArgArgGlyLeuGluAspAlaIleArgThrGlnAsnArgAspMetGluAsn 155
         :::||:::|||:::||||||
Db  929 TTG-----ACAATCCAGCCCTTCAAGATGAGCTGCCAACCCAGAGACCCTCAACCAC 982

Qy  156 ThrAspProSerGluGlnIleAlaLysAlaValMetAspLysPheGlnThrGlnIleIleLeu 175
```

```
Db  983 CTCCTCCAGCAAGAGAGTGGCACCGAGGAGCG---GAGCACTTCACCATCGAGCTGACC 1039
         |||                           :::                    ::::::
Qy  176 ProGlyLeuValAlaAsnMetIleAlaGlyAlaGlyAlaAsnProPheLysMetProGlnMet 195
Db 1040 ------------------------GAGGAGAACTTTAGGCGGCGGGTTCCAGCCGCGAG 1069
                                  :::   |||:::|||
Qy  196 ArgLysAlaGlnAlaAlaAlaProSerValPheGlnAlaLeuAlaGlnArgAlaMet 215
Db 1070 CATGACAGGCAGGCTAAGGAGCTGTTCCTTTGGAAGAACATTAGAGAGAAATGGAGCTG 1129
         |||||                               :::  |||
Qy  216 LeuGlyLysAsnAlaProValAlaLysGlyGlyGluGlnArgMetMetMetAsn 235
Db 1130 ------------------------AGAATTGAAACGCAGAAACAAACCCTCAAT 1159
                                  :::|||                 ::::|||
Qy  236 ArgValAspArgMetGlnGln----------ArgGluLeuGlnGluGlnGlnGluAspAsp 253
Db 1160 GCCCGAGATGAGTCAATTAAAAACTTCTGAGATGTTGCAAAGTAAAGCTTGCCATCC 1219
         ||||||    :::::::::   |||   :::
Qy  254 AspAspLeuGluAspValProArgArgSerSerAspGlyGluProGln--- 272
Db 1220 AAAAGCCTGGAGGATGACAATGAGCGACGCGGCGATGGCGAGAGGCTCGAGTCTTCAGGTC 1279
         ||||:::|||:::                   ||||||::::::
Qy  273 -------------------------SerGluAlaGluHisGlnArgArgAsp 281
Db 1280 AGCCACTTGGAAGTGATTTTAGATCAGAAGAAGAAGAAAACATACATCTTAGAGAGGAA 1339
         :::
Qy  282 LeuAlaArgArgLeuLysSerSerProArgLeuLysGluLeuGlnAsnAlaGluVal 301
Db 1340 TTGCACCGGAAGAGCCAACTTCAGCCG       GAGCCAGCCAAGACG 1381
         |||:::      |||:::::
Qy  302 GlnSerLeuLeuSerTyrGlnArgMetArgAspSerProLeuSerLysArgArgProLeu 321
Db 1382 AAGGCTCTCCAGACTGTCATCGAAATGAAGGACACAAAAATCGCTTCA-------TTG 1432
         ::::::|||   ||||||::::::           ||
Qy  322 AlaMetAsnAspGluAsp---GluSerAlaPheArgAlaMetGlyAlaArgAlaLysLeu 340
Db 1433 GAACGAGAAACATAAGGGATCTTGAGGATGAGATCCAGATGTTAAAAGCCAATGGTGTGCTG 1492
         :::::::|||
Qy  341 AspGlnLysSerGlnLeuValLeuGlyPheGlyPheGluSerAspAspAspAspGlu 360
Db 1493 ------------------------AACACTGAG 1501
                                  ::: |||
Qy  361 AspGluAspGluAsnLeuIleLeuAspProSerGluAsnSerPheArgAlaProLeu 380
Db 1502 GACCGCGAAGAGAG- 1516
         ||:::||
Qy  381 ArgLeuSerSerGlyPheValGluLysGluLysLysSerAsnAspGluLeuLysSerAlaLeu 400
Db 1517 ------------------ATCAAACAAATTGAGGTTTACAAAGTTCACTCC 1549
                           :::  ||||||   ||||||
Qy  401 AspArgIleLysTyrArgValAspAspValGluLysProLysProMetGlu 420
Db 1550 AAGTTTATGAAGACCAAGATTGATCAGCTGAAGCAGGAGAACTTTCAAAGAAAGAGTCGGAA 1609
         |||:::||    |||||||||
Qy  421 PheAsnProLysProGlyThrPheAlaProArgLysIleProThrArgProArg 440
Db 1610 CTT-------------CTTGCCTTACAAACAAAGCTTGAAACCCTGCAGCAAT 1648
         :::
Qy  441 LysMetLeuProLeuLeuIleGlySerAspProLysValGlnGluGlyIleArgHis 460
Db 1649 CAA----------AATTCAGATTGCAAGCAACACATTGAGTGCTCAAAGAG 1690
         :::              ||||||   ||||||
Qy  461 ProSerThrTrpLysIleAlaLysGluSerArgValLeuThrAsnLeuLysAsnAsn 480
Db 1691 --------GCCAAGAACAGAGGGCTGCCATC--CTTCAGACTGAG 1735
         :::         |||:::|||
Qy  481 ProSerLeuAlaLeuPheMetLysAspAspLys---------LeuGluAsnThrLeuLys 497
Db 1736 GTAGATGCCTGAGATTACGACTGGAAGAAAAAGAAATCTTCCTCAATAAAAAAACAAAA 1795
         |||:::::
Qy  498 GlyArgGlnMetLeuThrAspGluGlnLysGlyLysArgThrArgIleArgValAla 517
```

```
Db   1796 CAGCTCACGAGGACCTCACA---GAAGAGAGGGG------------  1825
                |||:|||                                        Qy   46 LeuAlaProSerMetGluAlaLeuGlu------------------------  54
                |||:|||      :::|||::::|||                      Db 3979 CTGAGCACCAGCTCAAGCAGTGGAGGACGAGAGAAGAATTCCTCCGGAGCAGCTGCGAG  4038

Qy  518 LeuProArgLeuPheGlyAlaProThrAlaLysAlaGluMetIleAspAlaLysValPhe  537
             |||:::                                            Qy   55 ----------------------------LeuMetGlyValGlnPhe---  60
Db 1826 -----------------ACACTGGCGCCGGTGAAATTCGTGACATGAAAGATATG  1861      Db 4039 GAGGAGGAGGGCCAGCACACACCTGGAGACAGCAGCAGCACCCTCCATGCCCAGTG  4098

Qy  538 GlnAspIleGluGluProProLeuPheGluProLysGlyArgHisThr  557
             |||:::::::||||||                                  Qy   61 ValAspAlaLeuIleLysLysGlyGlnMetGluMetAlaLysAlaLysGlyAlaLysPheLysThrGln  80
Db 1862 TTAGAAGCTGAAGGAAGAAAAATCAATGTTCTTCAGAAAAAGAATTGAAAACTTGCAAGAA  1921     Db 4099 ---------ATGAAAGAAGATGGAGCAGTGTGGGGTGCCTGGAAACTGCT  4149

Qy  558 ArgLeuArgTrpThrGlyAlaAsnGluLysGluLeuProGlyLeuGlySerArgPheIle  577
             :::|||||||                                        Qy   81 LeuGluValLeuLysValHisProAsp------  90
Db 1922 -----------GATAAAGACAAGCAACTGACCAACCTGAAAGACAGA  1966    Db 4150 GAGGAGTGAAGAAGAGGAAGGCCTGAGCGAGGCCACGAGGAG  4209

Qy  578 LeuProSerLeuAspProThrMetProAlaLeuAsnThrAlaPheSerThr------  594
             :::||||||                                          Qy   91 ------GlnPheAspLysTyrLysLysLeuLys----------ValAsp  101
Db 1967 GTGAAGTCCCTTGCAGACGGATTCCAGTAATACAGATACTGCACTGGCGACCTAGAGGAA  2026   Db 4210 AAGGTGCCGCCTACGACAAGCTGGACAAGACCGCGCTGAGCCAGCAGAGCTGAC  4269

Qy  595 ----------GlnGlyArgGlyArgAspGlu  601
             :::|||                                             Qy  102 AspLeuAlaAlaAspAlaValMetGlnGlnIleAlaGluGlnGlnLysLeuGlnMetAlaLysAlaLysLeuGlnProLysSer  121
Db 2027 GCTCTGTCGAGAGAGGAGATAATTGAGCGTGAAAGAACAGCGGAAGAGATGAT  2086    Db 4270 GACCTGCTGGTGACCTGGACCACCAGCGCGTGCAACCTGGAAGAAGCAG  4329

Qy  602 TrpAspThrMetPheLysIleProAsnAsnTrpAsnProGlyAspGluValGlyPheLys  621
             :::  :::|||                                        Qy  122 GlyAsnAlaPheIleAspMetLeuAsnGlyAsnGlyIleProIleGly---------  137
Db 2087 CGGGAAAGACTAGAAGAGATAGAATCCTTCCGAAAAGAGAACAACAGACCTGAAAGAGAAG  2146    Db 4330 AAGAAG------TTTGACCAGTCCTCGGCGAGGAGAAGACCATCTCTGCCAAGTATGCA  4383

Qy  622 MetAsnSer 624                                                       Qy  138 ----------SerSer 139
Db 2147 GTCAATGCT 2155                                                      Db 4384 GAGGAGGCGCGACCGGCTGAGCGCGAGGCCCGAGGAGGAGGAGGACCAAGGCTCTGTCGCTG  4443
```

RESULT 3
US-09-949-016-5001

```
                                                  Qy  140 IleArgGlyLeuGluAspAlaLeuAlaLeuArgThrGlnAsnThrArgAspMetGluAsnThrArgAspProSer  159
                                                         :::  :::||||||||                                    Db 4444 GCCCGGGCCCTGGAGGAAGCCATGACGCGGCAGAGGGCGGAGCTGGAGCGCGGCTCAAC---  4497

Qy  160 GluGlnIleAlaAlaLysAlaValMetAspLysProPheGlnThrGlnIleLeuProGlyLeuVal  179
             :::                                                Qy  ---ATG
             :::|||:::|||                                       Db 4498 ---------------------AAGCAGTTCCGACACGGAG-----  4518

Qy  180 AlaAsnMetIleAlaGlyLysGlyAsnProPhe---LysMetProGlnGlnMetArgLysAla  198
             |||:::||||||                                       Db 4519 GAGGACCTTATGCTCCAAGATGATGTGGCAAGAGTGTCCAGAGCTGGTGCACGAGAAGTC-  4577

Qy  199 GlnAlaAlaProSerValPheGlnLysLysGlnAlaLeuAlaLeuAlaGlnAlaMetLeuGlyLys  218
             |||                                                Db 4578 CAAGCGGGGCCCT-----------------------AGAGCAGCAGGTGGAGGA  4607

Qy  219 Asn------------AlaProValAlaGlyLysGlyAsnProGlnGlnLysArgMetMetAsn  235
             :::                                                Db 4608 GATGAAGAGCAGGCTGGAGAGCTGGAGGACCTGGAGGAGCTCCAGGCCACCGAAGATGCCAAGCT  4667

Qy  236 -ArgValAspLysAspArgMetGln-------------------GlnArgGluGlnLeuGluGl  249
             :::|||:::                                          Db 4668 GCGGTTGAGGTGCAACCTGCAGCCCATGAAGGCCAGTTCAGAGAGCCACCTTGAGCGGCGCGCCG  4727

Qy  249 uAspGlyAspAspAspAspAspLeuGluAspValProArgProArgArgProArgSerSerAspGl  269
             |||                                                Db 4728 GGACGAGCAGCAGGAGGAGGAGGAAGAGCAGCTGGTTCAGAGACGTGCGGGAGATGGAGGC  4787

Qy  269 yGluProGlnSerGluGlnAlaGluHisGlnAlaLeuArgArgLeuLysSerSe  289
             :::                                                Db 4788 AGAGCTGGAGGACGAGGAAGCAGCCTCGATGGACGTGGCTCGCGAAGAAGCTGGA  4847

Qy  289 rProArgLeuLysGluLeu---------------------------------  295
             ||||||:::|||                                        Qy  GlyPheCysIleAlaProLeuSer-------AlaGlnSerProSerThrSerAsp-------  27
             :::|||::::|||                                      Db 4848 GATGGACCTGAAGGACCTGGAGGCTCACGTGCACTCGGCTCAAGAACCGGGACGAAGC  4907

Qy  296 ---------LeuGlnAsnAlaGluValGlnSerLeuLeuSerTyrGlnArgMetArgAs  312
             ||||||                                             Qy   28 AlaProGly-----AlaLeuLeuSerSerLeuValGlyLysSerHisGlnLysLeuPro  45
             :::|||                                            Db 4908 CATCAAACAGCTGCGAAGCTGCAGCTCGCATGCGCGAGCTGGATGA  4967

Qy  312 pSerProLeuSerLysArgArgProLeuAlaMetAsnAspSerGluSerAlaPheAr  332
             :::|||||||                                         Qy   12 GlyPheCysIleAlaProLeuSer-------AlaGlnSerProSerThrSerAsp-----  27
Db 3860 GGCTTCTTCAGCCAGTCCGACAGCTCCAGCAAGCTTCACCAAGGACTTCTCCGCGCTGG  3919
             ||||||                                             Qy   28 AlaProGly-----AlaLeuLeuSerSerLeuValGlyLysSerHisGlnLysLeuPro  45
Db 3920 AGTCCCAGCTGCGAGGACCACTCAGGAGCTG-CTGCAGGAGGAGAACCGGACCAGAAGCTGAGC  3978
```

Alignment Scores:
Pred. No.:   2.54e-06   Length: 5883
Score: 162.50   Matches: 136
Percent Similarity: 35.90%   Conservative: 88
Best Local Similarity: 21.79%   Mismatches: 213
Query Match: 4.94%   Indels: 192
DB: 4   Gaps: 25

US-10-736-868-2 (1-643) x US-09-949-016-5001 (1-5883)

```
Db  4968  CACCCGGCGCTCTCGTGAGGAGATCGCCAGGCCCAAAGAACAGGAGAAGAAGCTGAA  5027
                |:::        ||:::|||               ||||||
Qy   332  gAlaMetGluAlaArgAlaLeuAspLysSerGlnLysValLeuValLeuHisGl     352
                :::|||||||||
Db  5028  GAGCATGGAGGCC------------GAGATGATCCAGTTGCGAGGAGGAACTGGCAGC  5072
Qy   352  yPheGlyGluSerAspAspAspAspGluAsnLeuIleAspProSerGl           372
                :::   |||:||:||  |||
Db  5073  CGCGGAGCGTGCCAAGCGCCAGCCCAGCGAGGCGGATGAGCTGGCTGACGAGATCGC  5132
Qy   372  uAsnSerPheArgArgAlaProLeuArgLeuSer------------             383
                ||||||  :::   |||
Db  5133  CAACAGCAGCCAGAAAGAGCCCTGGCTTAGAGGAGGAAGCGGCGTCTCGGAGCCCGCAT 5192
Qy   384  -SerGlyPheValGlyLeuLysSer------------------AsnAspGl         395
                :::||||||||||:::
Db  5193  CGCCCAGCTGGAGGAGGACTGGAGGAGGCAGGCAACACGAGCTGATCAACGACCG     5252
Qy   395  uLeuLysSerAlaLeuAspArgIleLysTyrArgValAspValAspGluLysTyrLeuAl 415
                ||||| :::||:|
Db  5253  GCTGAAGAAGGCC------------AACCTGCAGATCGACCAGATC---            5286
Qy   415  aProLysProMetGluPheAsnProLysProGlnTyrPheAlaProArgLysIl     435
                :::|||:||:|||||
Db  5287  ------------AACAC-CGACCTGAACCTGA--------                    5306
Qy   435  eProThrArgProArgLysMetLeuProLeuLeuIleGlySerAspProLysValGlnGl 455
                :::|||||:|||||
Db  5307  ---GCGCAGCCACGCCC-AGANGAACGAGAATGCTCGGCAGCAGCTCGGAACGCCAGAAC 5361
Qy   455  uGluIleLeuArgArgHisProSerThrGluTrpLysIleAlaLysGluSerArgValLeuTh 475
                |||:||  |||
Db  5362  AAGGAGCTTAAGGTCAAGCTGCGAGGAGATGG----------AGGGCACTGTC      5403
Qy   475  rAsnLeuLysAsnAsnProSerLeuAlaAlaLeu------------PheMetAspAs  490
                ::::  |||||
Db  5404  AAGTCC-AAGTACAAGGGCTCCATCACCGCCCCTCGAGGCCCAAGATTGCACAGCTGGACGA 5462
Qy   490  pLysLeuGluAsnThrLeuLysGlyArgGlnMetLeuThrAspGluGlnLysLysGlyArgTh 510
                :::||:::|||  |||||
Db  5463  GCAGCTGCAACAGCACGAGCAAGGAGCGCCAGGAGCCCTGC---AAACAGGTGCGTCGGAC 5519
Qy   510  rArgValLys  513
                |||||
Db  5520  CGAGAAGAAG  5529
```

FILING DATE: September 25, 1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Smith, DeAnn F.
REGISTRATION NUMBER: 36683
REFERENCE/DOCKET NUMBER: 2115-0086900B
TELECOMMUNICATION INFORMATION:
TELEPHONE: (810) 641-1600
TELEFAX: (810) 641-0270
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 2680 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA to mRNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
INDIVIDUAL ISOLATE: Sample 2
TISSUE TYPE: Acute myelomonocytic leukemia, M4Eo
TISSUE TYPE: subtype (inv16)
POSITION IN GENOME:
CHROMOSOME/SEGMENT: 16[inv(16) (p13q22)]
FEATURE:
NAME/KEY: CDS
LOCATION: 1..2451
US-08-533-306A-5

Alignment Scores:

Pred. No.:	1.73e-06	Length:	2680
Score:	159.00	Matches:	150
Percent Similarity:	38.30%	Conservative:	130
Best Local Similarity:	20.52%	Mismatches:	230
Query Match:	4.83%	Indels:	225
DB:	2	Gaps:	35

```
US-10-736-868-2 (1-643)  x  US-08-533-306A-5 (1-2680)

Qy   15  IleAlaProLeuSerAlaGlnSerProSerThrSerAspAlaProGlyAlaProGlyAlaLeuLeuSer  34
Db  562  GTGGCGTCCCTCAGTTCCCAGCTCCAGGACACCCAGGAGTTG--------------            603
Qy   35  SerLeuValGlyLysSerHisGlnLysGlyLeuProLeuAlaProSerMetGluAlaLeuGlu  54
Db  604  ---CTTCAAGAGACAAACCCGGCAGAAGCTCAACGTGTCTACGAAGCTGCGCCAGCTGGAG       660
              ||||| |||||||||||||||
Qy   54  -----  54
Db  661  GAGGAGGCGGAACAGCCTGCAAGACCAGCCTGGACGGAGGAGATGGAGGCCAAGCAAGAACCTG   720
Qy   55  -----LeuMetGlyValGlnPheValAspAlaLeuIleLysLysGlyGln  69
Db  721  GAGCGCCACATCTCCACTCTCAACATCCAGCTCTCCGACTCGAAG--AAGAAGCTGCAG        777
                 :::|||  |||||
Qy   70  ------MetGluMetAlaLysLysGlyAlaLysPheLysThrGlnLeu  81
Db  778  GACTTTGCCAGCACCGTGGAAGCTCTGGAAGCAGGGGAAGAAGAGGTTCCAGAAGGAGATC      837
Qy   82  GluValLeuGluLysValHisProAspGln------------PheAspLysTyrLysLysLeu  98
                ||| :::  :::  :::    :::|||
Db  838  GAGAACCTCACCCAGCAGTACGAGGAGGAAGGCGCCGCTTATGATAAACTGGAAAAGACC      897
Qy   99  Lys------------ValAspAlaLeuAlaAlaAspAlaValMetGlnAlaAla  112
                |||  :::   :::
Db  898  AAGAACAGGCTTCAGCAGCCTGGACGACCTGGTTGTTGATTTGGACACACCAGCGCAA        957
Qy  113  GluMetAlaLysLeuGlnProLysSerGlyAsnAlaPheIleAspMetLeuAsnGlyAsn  132
                :::  |||  :::
Db  958  CTCGTGTCCAACCTGGAAAAGAAGCAGCAGGAAA------TTTGATCAGTTGTTAGCCGAG    1011
Qy  133  GlyIleProIleGlySerSer-------------------------  139
```

RESULT 4
US-08-533-306A-5
; Sequence 5, Application US/08533306A
; Patent No. 5837457
GENERAL INFORMATION:
APPLICANT: Liu, Pu
APPLICANT: Collins, Francis S.
APPLICANT: Siciliano, Michael J.
APPLICANT: Claxton, David
TITLE OF INVENTION: Markers for Detection of Chromosome 16
TITLE OF INVENTION: Rearrangements
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: Harness, Dickey & Pierce, P.L.C.
STREET: P.O. Box 828
CITY: Bloomfield Hills
STATE: MI
COUNTRY: USA
ZIP: 48303
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/533,306A

```
Db  1012  GAGAAAAACATTCTTCCAAATACGCGGATGAGAGGGACCAGAGCTCGAGCGCAGAAGCCAGG  1071
             |||              |||
Qy   140  --------------IleArgGlyLeuGluAspAlaIleArgThrGln  150
                            |||||:||:
Db  1072  GAGAAGCGGAAACAAGGCCCTCGGCCCTTGGGGCCCTTGAAGAGGCCTTGGAAGCCAAA  1131
Qy   151  ArgAspMetGluAsnThrAspProSerGluGlnIleAlaLysAlaValMetAspLysPhe  170
          :::||:|||            :||||:
Db  1132  GAGGAACTCGAGCGGACCAAC--------------AAAATGCTCAAAGCCGAAATGGAA  1176
Qy   171  GlnThrGlnIleLeuProGlyLeuValAlaAsnMetIleAlaGlyIleLysAsnProPhe---  189
             |||                        |||||:
Db  1177  --------------GACCTGGTCAGCTCCAGGATCAGCGGATGACGTGGGC  1206
Qy   190  LysMetProGlnGlnMetArgArgLysAlaGlnAlaAlaProSerSerValPheGlnGlnAla  209
          :::::                |||          :::||:
Db  1207  AAGAACAGAGTCCATGGACGAAGAAGTCCAAGCGGGCCCCTGGAGACCCAGATGGAGGAGATG  1266
Qy   210  LeuAlaGlnLysArgAlaMetLeuGlyLysLysAsnAlaProValAlaProValAlaGlyArgGlyGluGlu  229
             |||                      :::                  |||||:::
Db  1267  AAGACGCCAGCTGGAAGACCTGGAAGACGAGCTGCAAGCCTCG-----------GAGGAC  1314
Qy   230  GlnArgMetMetMetAsnArgValAspGlyValAspGlyMetGln-----------------GlnGln  243
          :::                                |||
Db  1315  GCCAAACTG-----------CGGCTGGAAGTCAACATGCAGCGCGCGCCTCCAAGGGCCGCCAGTTCGAA  1365
Qy   244  ArgGluLeuGlnGluGlnGluGluAspGluAspSerProArgLeuGluAspValProArg  263
          :::::::::::::::              :::
Db  1366  AGGGATCTTCCAAGCCCCGGACCCCCGGGACGAGCAGAAATGAGGAG-----------AAG  1404
Qy   264  ArgArgSerSerAsp-----------GlyGluProGlnSerGluAlaGluHisGlnAlaArgArg  280
          |||||                   |||          :::||:
Db  1405  AGGAGGCAACTTGCAGAGACAGCTTCACGAGTATGAGAACGGAACGAGGACAGCGAAAC  1464
Qy   281  AspLeuAlaArgArgLeuLysSerSerProArgLeuGluLeuLeuGlnGlnAsnAlaGlu  300
          :::                    |||||:::
Db  1465  GAACGTGCCCTGGCAGCTGCGAGCAGAAGAGAAGCTGAAGGGGGACCTGAAAGACCTGGAG  1524
Qy   301  ValGlnSerLeuLeuSerThrTyrGlnAsnArgMetArgAspSerProLeuSerLysArgArgPro  320
             :::                  :::          |||||:
Db  1525  CTTCAGGCC-----------GACTCTGCCATCAAGGGAGGGGAGGAAGCAGCAGCTACGCAAA  1581
Qy   321  Leu-----------AlaMetAsnAsp-----------GluAspGluSerAlaPheArg----  332
          :::||:::                     |||
Db  1582  CTGCAGGCTCAGATGAAGGACTTTCAAAGAGAGCTGCAAGAGATGCCCGTGCCTCCAGAGAT  1641
Qy   333  AlaMetGluAlaAlaLysAlaLysAspLeuAspProGlnGlnLeuValLeuLeuGlyLeuHisGly  352
          |||                      :::
Db  1642  GAGATCTTTGCCACAGCCAAAGAGAATGAGAAGAGAAAGCCAAGAGCTTG-----------  1689
Qy   353  PheGlyGluSerAspAspSerProGluAspGluAsnLeuIleAspProSerGlu  372
          :::            |||||:::          |||
Db  1690  -----------GAAGCAGAGACCTCATGCAGCTACAAGGAGAC-----------CTCGCCGCCGCCTGAG  1734
Qy   373  AsnSerPheArgArgAlaProLeu-----------ArgLeuSerSerGlyPheValGluValLys  389
          :::                      |||:::
Db  1735  AGGGCTCGCAAACAAGCGGACCTCCGGAAGAGGAACGGAACTGGCCAGAGCTTGGCCAGTAGC  1794
Qy   390  LeuLysSerAsnAspGluLeuLeuSerSerAlaLeuAspArgIleLysTyrArgValAspAsp  409
             |||                      :::          |||
Db  1795  CTGTCCGGAAGGAACGACCACTCCAGAGAAGCGCCCTGGAGGCCCGATCGCCCAG  1854
Qy   410  ValGluLysTyrLeuAlaProLysProMetGluPheGlnProLysProGlnTyr  429
          ::::::          |||          :::||:
Db  1855  CTGGAGGAGGATGGAGGAGAATGGAGGCCATGGAGCGACCGGGTCC-  1912
Qy   430  PheAlaProArgLysIleProThr-----------Arg  438
             |||                      :::
Db  1913  GCAAAGCCACACAGAGCCCGAGCCAGCTGCCACAACGAGCTGCCACAGAGCCAGCACG  1972
Qy   439  ProArgLysMet---LeuProLeuLeuLeuGlySerAspProLysSerAspProLysValGlnGluGluIle  457
             |||:::                      :::
```

```
Db  1973  CCCAGAGAATGAGAGTGCCC--------GGCAGCAGCTCGAGCGCAGAACAAGGAGC  2023
             |||
Qy   458  ArgArgHisProSerThrGluTrpLysIleGluSerArgValLeuThrAsnLeu  477
Db  2024  TCCGGAGCAAGCTCCACGAGATGGAGGGG-GCCGTCAAGTCC--------  2064
Qy   478  LysAsnAsnProSerLeuAlaAlaLeu----------PheMetAspLysLeu  492
          |||                      :::||||||||:|||
Db  2065  AAGTTCAAGTCCACCATCGCGCGCGCCAAGATTGGAGGAGCAGGTC  2124
Qy   493  GluAsnThrLeuLysGlyArgGlnMetLeuThr  503
             |||
Db  2125  GAGCAGGAGGCCAGAGAAGAACAGGCCAGCCCAAGTCGCTGAAGCAGAAAAGACAAGAAG  2184
Qy   504  --------------AspGlu  505
                            |||
Db  2185  CTGAAGGAAATCTTGCTGCAGGTGGAGGACGAGCGCCAGAGATGGCCAGCAGTACAAGGAG  2244
Qy   506  ------GlnLysGlyArgThrArgValLysThrIleArgAlaLeuProArgLeuPheGly  523
                :::                      |||||:::
Db  2245  CAGGCAGAGAAGGCAATGCCAGGGTCAAGCAGCTCAAG--------AGGCAGCTGGAG  2295
Qy   524  AlaProThrAlaLysAlaGluMetIleAspAla--------------LysValPheGlnAspIle  540
             |||                      :::
Db  2296  GAGGCAGAGGAGGCGAGTCCCAGCGCATCAACGCCCCAAGTCGCAGGGGGGACTG  2355
Qy   541  GluGluArgProIleProLeuProPhePheGluProLysGlyProLysGlyArgHisThrArgLeuArg  560
             :::                      |||
Db  2356  GATGAGGCC  2364
Qy   561  TrpThrGlyAlaAsnGlu-----------LysGluIleProLeuGlySerArgPheIle  577
          :::||:::                   |||
Db  2365  ---ACGGAGCAACGAGCGGCCATGGGCCGTGGAGGTGAACGACTCAAGCTCAGA  2421
Qy   578  LeuProSerLeuAspProThrMetProAlaLeuAsnThrAlaPheSerThrGlnGlyArg  597
             |||                      :::
Db  2422  GGGCCCCCC--------CCACAGGA-AACTTCGCAGTGATGCACCAGGCGGAGG  2465
Qy   598  AlaArgAspGluTrpAspThrMetPheLysIle  608
             |||
Db  2466  AAACGAGACCTCTTCGTTCCTTCTTTAGAAGTC  2498
```

RESULT 5
US-08-742-923A-5
; Sequence 5, Application US/08742923A
; Patent No. 5869611
GENERAL INFORMATION:
; APPLICANT: Liu, Pu
; APPLICANT: Collins, Francis S.
; APPLICANT: Siciliano, Michael J.
; APPLICANT: Claxton, David
; TITLE OF INVENTION: Markers for Detection of Chromosome 16
; TITLE OF INVENTION: Rearrangements
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Harness, Dickey & Pierce, P.L.C.
; STREET: P.O. Box 828
; CITY: Bloomfield Hills
; STATE: MI
; COUNTRY: USA
; ZIP: 48303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/742,923A
; FILING DATE: No. 5869611ember 1, 1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Smith, DeAnn F.
; REGISTRATION NUMBER: 36683

REFERENCE/DOCKET NUMBER: 2115-00869DVC
TELECOMMUNICATION INFORMATION:
  TELEPHONE: (810) 641-1600
  TELEFAX: (810) 641-0270
INFORMATION FOR SEQ ID NO: 5:
  SEQUENCE CHARACTERISTICS:
    LENGTH: 2680 base pairs
    TYPE: nucleic acid
    STRANDEDNESS: double
    TOPOLOGY: linear
  MOLECULE TYPE: cDNA to mRNA
  HYPOTHETICAL: NO
  ANTI-SENSE: NO
  ORGANISM SOURCE:
    ORGANISM: Homo sapiens
    INDIVIDUAL ISOLATE: Sample 2
    TISSUE TYPE: Acute myelomonocytic leukemia, M4Eo
    TISSUE TYPE: subtype (inv16)
  POSITION IN GENOME:
    CHROMOSOME/SEGMENT: 16[inv(16) (p13q22)]
  FEATURE:
    NAME/KEY: CDS
    LOCATION: 1..2451
US-08-742-923A-5

Alignment Scores:
Pred. No.:
Score: 1.73e-06   Length: 2680
Percent Similarity: 159.00   Matches: 150
Best Local Similarity: 38.30%   Conservative: 130
Query Match: 20.52%   Mismatches: 230
DB: 4.83%   Indels: 225
                                      Gaps: 2

US-10-736-868-2 (1-643) x US-08-742-923A-5 (1-2680)

Qy   15 IleAlaProLeuSerAlaGlnSerProSerThrSerAspAlaProGlyAlaAlaLeuLeuSer   34
           :::||||||||||||||:::||||||
Db 562 GTGGGCTCCCTCAGTTCCCAGGTCACCCAGGAGTTG----------------------- 603

Qy   35 SerLeuValGlyLysSerHisGlnLysLeuProLeuAlaProSerMetGluAlaLeuGlu   54
          ||| :::::: |||||||| :::::: |||||||
Db 604 --CTTCAAGAGAAACCCGGCGAAGCTCAACGTGTCTACGAAGCTGCGCCAGCTGGAG   660

Qy   54 ...Ser   54
Db 661 GAGGAGCGGACAGCAGCCTGCCGAGACAGCAGCTGGACAGGAGGAGATGGAGGCCAAGCAGAACCTG   720

Qy   55 ---LeuMetGlyValGlnIlePheValAlaAspAlaLeuLeuIleLeuLysLysGlyGln   69
             :::||||||||||||||:::|||
Db 721 GAGCGCCACATCTCAACATCCAGCTCTCCGACTCGAAG--AAGAAGCTGCAG   777

Qy   70 ---MetGluMetAlaLysSerGlyAlaAsnPheLysThrGlnLeu   81
             :::|| ||| ::::::
Db 778 GACTTTGCCAGCACCGTGGAAGCTCTGGAAGAGGGAAGAGAGGTTCCAGAAGGAGATC   837

Qy   82 GluValLeuGluLysValHisProAspGln-------PheAspLysTyrLysLysLeu   98
          ::|| |||||| ||| ::::::
Db 838 GAGAAGACCGGACAGCAGTACGAGGAGGACCGCGGCCCGCCCCTTATGGTAGACTGAAAAGACC   897

Qy   99 Lys------ValAspAspSerLeuAlaAlaAspAlaAspAlaValAlaMetGlnGlnAla   112
          ||| ::|||||||||
Db 898 AAGAAGAGGCTTCAGCAGGAGCTGGACGACCTGGTTGTTGATTGGACAACCAGCCGGCAA   957

Qy 113 GluMetAlaLysLeuGlnLysProLysSerArgGlyAsnAlaPheIleAspAspMetLeuAsnGlyAsn   132
          ::::: |||||::::: ||| |||
Db 958 CTCGTGTCCAACCTGGAAAAGAGAGCAGAGGAAA------TTTGATCAGTTGTTAGCCGAG   1011

Qy 133 GlyIleProIleGlySerSer-------   139
          |||||||:::::
Db 1012 GAGAAAAACATCTCTTCCAAATACGCCGATGAGAGGGACAGAGCTGGAGGAGAAGCCAGG   1071

Qy 140 -------IleArgGlyLeuGluAspAlaIleArgThrGln   150
                 ||| ||||||:::||||
Db 1072 GAGAGGAAACCAAGCCCTGTCCCTGGCTCGGCGCCTTGAAGAGGCCTTGGAAGCCAAA   1131

Qy 151 ArgAspMetGluAsnThrAspProSerGluGlnIleAlaAlaValAlaLysAlaValAlaMetAspLysPhe   170
          |||||:::||| :::|||
Db 1132 GAGGAACTCGAGCGGACCAAC-------AAAATGCTCAAAGCCGAAATGGAA-------   1176

Qy 171 GlnThrGlnIleLeuProGlyLeuValAlaAlaAsnMetIleAlaGlyLysAsnProPhe---   189
          :::::::::::::::::::::::::
Db 1177 -------------------------GACCTGGTCAGCTCCAAGGATGACGTGGGC   1206

Qy 190 LysMetProGlnMetArgLysValAlaGlnIleAlaAlaAlaProSerValPheGlnIleAlaAla   209
          ||:::::|||:::||||||
Db 1207 AAGAACGTCCATGAGCTGGAGAGTCCAACGGCCCTCGAGACCCAGATGGAGGAGATG   1266

Qy 210 LeuAlaGlnIleArgAlaMetLeuGlyLysAsnAlaAlaProValAlaGlyLysGlyArgGlyGluGlu   229
          |||||| ::: ||||||
Db 1267 AAGACAGCTGGAAGAGCTGGACGACAGAGCTGCAAGCCTCG---------GAGGAC   1314

Qy 230 GlnArgMetMetMetAsnArgValAspGlnArgMetGln--------Gln   243
          :::::::::: |||||||
Db 1315 GCCAAACTG--------CGGCTGGAAGTCAACATGCAGCCGCCTCAAGGGCCAGTTCGAA   1365

Qy 244 ArgGluLeuGlnLeuGlnAspGluAspAspAspLeuGluLeuGluHisGluValProArg   263
          ::: ::: |||
Db 1366 AGGGATCTTCCAAGCCCGGGACGAGCAGAATGAGGAG-----------------AAG   1404

Qy 264 ArgArgSerSerAsp---------GlyGluProGlnSerGluAlaGlnHisGlnArgArg   280
          |||||| :::||
Db 1405 AGGAGGCAACTGCAGAGACAGCAGTTCACCAGGATGAGACGGGAGGAGGAGGAGAAAC   1464

Qy 281 AspLeuAlaArgArgLeuLeuSerSerProArgLeuLysGlyLeuLeuGlnAsnAlaGlu   300
          ::: |||||| :::::: |||
Db 1465 GAACGTGCCCTGCCAGCTGCAGCAAAGAGAAGAGCTGGAAGGGGACCTGAAAGACCTGGAG   1524

Qy 301 ValGlnSerLeuLeuSerTyrGlnArgMetArgAspSerProLeuSerSerArgArgPro   320
          ::: |||
Db 1525 CTTCAGGCC---GACTCTGCCATCAAGGAGGGGAGGAGGAGCCATCAAGCACTACGCAAA   1581

Qy 321 Leu------AlaMetAsnAsp-------GluAspGluSerAlaPheArg---   332
          ||| :::|||
Db 1582 CTGCAGGCTCAGATGAACGACTTTCAAAGAGACGTGGAAGATGCCCGTGCCTCCAGAGAT   1641

Qy 333 AlaMetGluAlaArgAlaAlaLysLeuAspLysGlnLysSerGlnLeuValLeuGlyLeuHisGly   352
          ::: |||||| :::::: |||
Db 1642 GAGATCTTTGCCACAGCCAAAGAGAATGAGAAGAAAGCCAAGAGCTTG-----------   1689

Qy 353 PheGlyGluSerAspAspAspAspGluGluAspGluGluAspAsnLeuIleAspProSerGlu   372
          ::: |||
Db 1690 ------GAGGCAGACCTCATGCAGCTACAGAGGAC------------CTCCGCGCCCTGAG   1734

Qy 373 AsnSerPheArgArgAlaProLeu-------ArgLeuSerSerGlyPheValGluLys   389
          :::|||
Db 1735 AGGGCTCGGCAAACAAGCCGGACCTCGAGAAGGAGGACGAACTGGCAGGAGGAGCTGGCCAGTAGC   1794

Qy 390 LeuLysSerAsnAspGluLeuLysSerSerAlaLeuAspLysArgIleLysTyrArgValAspAsp   409
          ||:::::|||||:::::::
Db 1795 CTGTCGGGAAAGGAACCACTCCAGGACGAGAGGCGCAGCCCGGGATCGCCCAG   1854

Qy 410 ValGluLysTyrLeuAlaProLysProMetGluProAsnProLysPheAsnProGlnProGlyTyr---   429
          ::: |||
Db 1855 CTGGAGGAGGAGCTGGAGGAGGCAGGG-CAACATGAGGGCCATGAGCGACCCGGTCC---   1912

Qy 430 PheAlaProArgLysIleProThr-----------------------------Arg   438
          |||
Db 1913 GCAAAGCCACACAGCCGGCCAGCTCAGCAACGAGCGTGGCCACAGCGCGACACGG   1972

Qy 439 ProArgLysMet---LeuProLeuLeuGlySerAspArgProLysValGlnGluGluIle   457
          |||:::::: |||
Db 1973 CCCAAGAAGAATGAGAGTGCCC----------GCCAGCGCAGTCGAGCGGCAGAACAAGGAGC   2023

Qy 458 ArgArgHisProSerThrGluTrpLysTrpIleAlaLysGluSerArgValLeuThrAsnLeu   477
          ::: :::||||||
Db 2024 TCCGGAGCAAGCTCCACAGAGATGAGGGG-GCCGTCAAGTCC----------------   2064

```
SEQUENCE CHARACTERISTICS:
    LENGTH: 2887 base pairs
    TYPE: nucleic acid
    STRANDEDNESS: double
    TOPOLOGY: linear
    MOLECULE TYPE: cDNA to mRNA
    HYPOTHETICAL: NO
    ANTI-SENSE: NO
    ORIGINAL SOURCE:
        ORGANISM: Homo sapiens
        INDIVIDUAL ISOLATE: Sample 1
        TISSUE TYPE: Acute myelomonocytic leukemia, M4Eo
        TISSUE TYPE: subtype (inv16)
    POSITION IN GENOME:
        CHROMOSOME/SEGMENT: 16[inv(16) [p13q22]]
    FEATURE:
        NAME/KEY: CDS
        LOCATION: 1..2658
US-08-533-306A-3
```

```
Alignment Scores:
    Pred. No.:               1.95e-06
    Score:                   159.00
    Percent Similarity:      38.30%
    Best Local Similarity:   20.52%
    Query Match:             4.83%
    DB:                      2

                Length:        2887
                Matches:       150
                Conservative:  130
                Mismatches:    230
                Indels:        225
                Gaps:          35
```

US-10-736-868-2 (1-643) x US-08-533-306A-3 (1-2887)

```
Qy    15  IleAlaProLeuSerAlaGlnSerProSerThrSerAspAlaProGlyAlaLeuLeuSer  34
              :::|||||||              |||:|
Db   769  GTGGCGTCCCTCAGTTCCCAGCTCCAGGACACCCAGGAGTTG------------------  810

Qy    35  SerLeuValGlyLysSerHisGlnLysLeuProLeuAlaProSerMetGluAlaLeuGlu  54
              |||                |||||||       ::::                :::||
Db   811  ---CTTCAAGAGAACCCGGCAGAAGCTCAACGTGTCTACGAAGCTGCGCCAGCTGGAG   867

Qy    54  ------------------------------------------------------------  54
Db   868  GAGGAGGCGGAACAGCTGCAAGACGCTGCAAGCGGAGATGGGAGGCCAAGCAGAGACCTG  927

Qy    55  -----------LeuMetGlyValGlnPheValAspAlaLeuArgLeuIleLysLysGlyGln  69
                        ::|||      :::|||       |||||
Db   928  GAGCGCGCCACATCCACTCCAACATCCAGTCTCCGACTCGAAG---AAGAGCTGCAG   984

Qy    70  -------------------------MetGluMetAlaLysGlyAlaPheLysThrGlnLeu  81
                                         ::|||       ::|||      ::::
Db   985  GACTTTGCCAGCACCGTGGAAGCTCTGGAAGAGGGGGAAGAGAGGTTCCAGGAAGGAGATC 1044

Qy    82  GluValLeuGluLysValValHisProAspGln-----------PheAspLysTyrLysLeu  98
              ||:||      |||       |||            :||||||      ::||||
Db  1045  GAGAACCTCACCCAGCTACGAGGAGAAGGGCGGCCGCCTTATGATAAACTGGAAAAGACC  1104

Qy    99  Lys------------ValAspAspLeuAlaAlaAspAlaValMetGlnInAla  112
               |||                   ::|||      |||     :||
Db  1105  AAGAACAGGCTTCAGGAGAGCTTGGACGACCTGGTTGTTGATTTGGACAACCAGCCGCCAA 1164

Qy   113  GluMetAlaLysLeuGlnLnProLysSerGlyAsnAlaPheIleAspMetLeuAsnGlyAsn  132
              |||||                  :||||       ::::
Db  1165  CTCGTGTCCAACCTGGAAAAAGAAGCAGAGGAAA------TTTGATCAGTTGTTAGCCGAG 1218

Qy   133  GlyIleProGlyGlySer------------------------------------------  139
              ::|||
Db  1219  GAGAAAAACATCTCTTCCAAATACGCGATGAGAGGACAGAGCTGAGGCAGAAGCCAGG   1278

Qy   140  -----------------------IleArgGlyLeuGluAspAlaIleArgThrGln  150
                                        |||||:|||       ::||:|||||      :::
Db  1279  GAGAAGGAAACCAAGGCCCTTCCCTGCTCCTGGCGGCCCTTGAAGCCAAA         1338

Qy   151  ArgAspMetGluAsnThrAspProSerGluGlnIleAlaLysAlaValMetAspLysLysPhe 170
              |||::||||       |||:::       :::||::::
Db  1339  GAGGAACTTGCGAGGCGGACCAAC----------AAAATGTCAAAGCCGAAATGGGA----- 1383
```

```
Qy   478  LysAsnAsnProSerLeuAlaAlaLeu------------PheMetAspAspLysLysLeu  492
              ::::::|||||||
Db  2065  AAGTTCAAGTCCACCATGCGCGGCCGCTGGAGGCCAAGATTGCACAGCTGGAGGAGCAGGTC 2124

Qy   493  GluAsnThrLeuLysGlyValArgGlnMetLeuThr---------              503
              :::|||       |||
Db  2125  GAGCAGGAGGCCAGAGAGAAACAGGCAAAACAGGCCACCAAGTCGCTGAAGCAGAAGACAAGAAG 2184

Qy   504  -----AspGlu-------                                           505
                    |||||||
Db  2185  CTGAAGGAAATCTTGCTGCCAGGTGAGGACCAGCGCAAGATGGCCGAGCAGTACAAGGAG 2244

Qy   506  -------GlnLysGlyGlyAlaThrArgValLysThrLysAlaLeuProArgLeuPheGly  523
                      :::|||||||            |||||||       :::::
Db  2245  CAGGCAGAAAGGCAATGCCAGGTCAAGACAGCTCAAG--------AGGCAGCTGGAG  2295

Qy   524  AlaProThrAlaLysGluMetIleAspAla----------LysValPheGlnInAspIle  540
              :::|||||||            |||:::
Db  2296  GAGGCAGGAGGAGTCCCAGCGCCATCAACCGCCAACGCTGCAGCCGGGAGCTG  2355

Qy   541  GluGluArgProIleProProLeuPhePheGluProLysGlyArgHisThrArgLeuArg  560
              :::|||
Db  2356  GATGAGGCC                                                    2364

Qy   561  TrpThrGlyAlaAsnGlu--------------LysGluIleProGlyLeuGlySerArgPheIle 577
              |||  :::|||||||
Db  2365  ---ACGGAGCAACGAGGCCATGGGCCGTGAGGTGAACCGCCACTCCAAGAGCAAGCTCAGA 2421

Qy   578  LeuProSerLeuAspProThrMetProAlaLeuAsnThrAlaPheSerThrGlnGlyArg  597
                                               ::::
Db  2422  GGGCCCCCC------------------------CCACAGGA-AACTTCGCAGTGATGCAGCTGGAGAGG 2465

Qy   598  AlaArgAspGluTrpAspThrMetPheLysIle  608
              :::        ::
Db  2466  AAACAGGAGACCTCTTTCGTTCCTTCTAGAAGGTC 2498
```

RESULT 6
US-08-533-306A-3
; Sequence 3, Application US/08533306A
; Patent No. 5837457
; GENERAL INFORMATION:
;   APPLICANT: Liu, Pu
;   APPLICANT: Collins, Francis S.
;   APPLICANT: Siciliano, Michael J.
;   APPLICANT: Claxton, David
;   TITLE OF INVENTION: Markers for Detection of Chromosome 16
;   TITLE OF INVENTION: Rearrangements
;   NUMBER OF SEQUENCES: 14
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Harness, Dickey & Pierce, P.L.C.
;   STREET: P.O. Box 828
;   CITY: Bloomfield Hills
;   STATE: MI
;   COUNTRY: USA
;   ZIP: 48303
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/533,306A
;   FILING DATE: September 25, 1995
;   CLASSIFICATION: 435
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Smith, DeAnn F.
;   REGISTRATION NUMBER: 36683
;   REFERENCE/DOCKET NUMBER: 2115-00869COB
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (810) 641-1600
;   TELEFAX: (810) 641-0270
; INFORMATION FOR SEQ ID NO: 3:

Qy   171 GlnThrGlnIleLeuProGlyLeuValAlaAsnMetIleAlaGlyLysAsnProPhe--- 189
                             ::::::::::||||| :::::
Db 1384 --------------------GACCTGGTCAGCTCCAAGGATGACGTGGGC 1413

Qy   190 LysMetProGlnMetLysArgAlaAlaGlnAlaAlaProSerValPheGlnGlnAla 209
         ||| :::::::::||||| :::   ::::::
Db 1414 AAGAACGTCCATGAGCTGGAGAGTCCAAGCGCGGCCCTGGAGACCCAGATGGAGGAGATG 1473

Qy   210 LeuAlaLeuGlnArgAlaMetLeuGlyLysAsnAlaProValAlaGlyArgGlyGluGlu 229
         ||| :::||| ||||||
Db 1474 AAGAGCGCAGCTGGAGAGAGCTGGAGGACGCAGCTGCAAGCCTCG-----GAGGAC 1521

Qy   230 GlnArgMetMetMetAsnArgValAspGlnMetGln--------------Gln 243
                    :::::||| :::
Db 1522 GCCAAACTG-----------CGGCTGGAAGTCAACATGCAGGCGCCTCAAGGGCCAGTTCGAA 1572

Qy   244 ArgGluLeuGlnGluGluGlnAlaGluAspAspAspLeuGluAspGluAspGluAspArg 263
         ::::::::||||||| :::
Db 1573 AGGGATCTTCCAAGCGCCGGGACGAGCAGAATGAGGAG-----------AAG 1611

Qy   264 ArgArgSerSer--------GlyGluProGlnSerGluAlaGluHisGlnArgArg 280
         ||||||
Db 1612 AGGAGGCACTGCGAGAGACAGCGTTCACGAGTATGAGACGGAACGTGAAGACGAGCGAAAC 1671

Qy   281 AspLeuAlaArgArgLeuLysSerSerProArgLeuLysGluLeuGlnAsnAlaGlu 300
         :::||| ::::
Db 1672 GAACGTGCCCTGGCAGCTGCAGCAAAGAAGAAGCTGGAAGGGGACCTGAAAGACCTGGAG 1731

Qy   301 ValGlnSerLeuLeuSerTyrGlnAsnArgMetArgAspSerProLeuSerLysArgArgPro 320
         :::::||| :::   :::
Db 1732 CTTCAGGGC--GACTCTTGCCATCAAGGGGAGGGAGAAGCCATCAAGCAGCTACGCAAA 1788

Qy   321 Leu--------AlaMetAsnAsp-------------GluGluSerAlaPheArg-- 332
         |||
Db 1789 CTGCAGGCTCAGATGAAGGACTTTCAAAGAGAGCTGGAAGATGCCCGTGCCTCCAGAGAT 1848

Qy   333 AlaMetGluAlaAlaArgAlaLysLeuAspLysArgGlnLysSerGlnLnLeuValGlyLeuHisGly 352
         :::||||||| ||||||
Db 1849 GAGATCTTTGCCACAGCCAAGAGAATGAGAAGAAGCCAAGAGCTTG------------- 1896

Qy   353 PheGlyGluSerArgAspAspAspGluAspGluAsnLeuIleAspProSerGlu 372
         ::::::||||||:::   :::   ::::||
Db 1897 -----GAAGCAGACCTCATGCAGCTACAAGAGAC--CTCGCCCGCCGCTGAG 1941

Qy   373 AsnSerPheArgArgAlaProLeu---------ArgLeuSerSerGlyPheValGluLys 389
         :::   :::||| |||
Db 1942 AGGGCTCGCAAACAAGCGGACCTGCGAGAAGGAGCAACTGGCAGGAGGCTGGCCAGTAGC 2001

Qy   390 LeuLysSerAsnAspGluLeuLysSerArgAlaLeuAspArgArgIleLysTyrArgValAspAsp 409
         ||| :::||:::
Db 2002 CTGTCGGGGAAGGAAACGCACTCCAGGAGGACCAGAAGCGCCGCCCTGGAGGCCCCGGATCGCCCAG 2061

Qy   410 ValGluLysTyrLeuValAlaProLysProMetGluPheAsnProLysProGlyProGlnTyr 429
         ::::|| :::   ::::::||||||
Db 2062 CTGGAGGAGGAGCTGGAGGAGGACAGCG-CAACATGAGGCCCATGAGCGACCGGGTCC- 2119

Qy   430 PheAlaProArgLysIleProThr------------Arg 438
         ||| :::|||
Db 2120 GCAAAGCCACACACAGCAGGCGCAGCTCAGCACGACGCTGGCCACGAGCGGCAGCACGG 2179

Qy   439 ProArgLysMet---LeuProLeuLeuLeuIleGlySerAspProLysValGlnGluGlnIle 457
         ||| :::|||
Db 2180 CCCAGAAGAATGAGAGTGCCC--------GGCAGCAGCTCGAGCGGCAGAACAAGGAGC 2230

Qy   458 ArgArgHisProSerThrGluThrProLysIleAlaLysSerArgValLeuThrAsnLeu 477
         :::::||||||:::   :::|||
Db 2231 TCCGGAGCAAGCTCCACGAGATGGAGGGG-GCCGTCAAGTCC-------------- 2271

Qy   478 LysAsnAsnProSerLeuAlaAlaLeu----------PheMetAspAspLysLysLeu 492
         ::::||||||:::||||||
Db 2272 AAGTTCAAGTCCACCATCGCGCGGCCGCTGGAGGCCCAAGATTGCACAGCTCGAGGAGCAGTC 2331

Qy   493 GluAsnThrGlnLysGlyArgGlnMetLeuThr----------------- 503
         ||| ::: :::|||
Db 2332 GAGCAGGAGGCCAGAGAGAAAAACAGGCAGCGACCAGTCGCTGAAGCAGAAAGACAAGAAG 2391

Qy   504 ----------------------AspGlu----- 505
                               |||||
Db 2392 CTGAAGGAAATCTTGCTGCAGGTGGACGAGCGCCAAGATGGCGCCAAGCAGTACAAGGAG 2451

Qy   506 ------GlnLysGlyArgThrArgValValThrIleArgAlaLeuProArgLeuPheGly 523
               :::|||||||
Db 2452 CAGGCAGAGAAAGGCAATGCCAGGTCCAGGGTCAAGCAGCTCAAG------AGGCAGCTGGAG 2502

Qy   524 AlaProThrAlaLysValAlaGluMetIleAspAla---------LysValPheGlnAspIle 540
         :::::::|||||| :::::::::
Db 2503 GAGGCAGGAGGAGGAGTCCCAGCCCATCAACCGCAGCAGCGAGCTGCAGCGGGAGCTG 2562

Qy   541 GluGluArgProIleProProLeuPhePheGluProLysProGlyHisThrArgLeuArg 560
         ::::||
Db 2563 GATGAGGGC-- 2571

Qy   561 TrpThrGlyAlaAsnGlu-------LysGlnIleProGlyLeuGlySerArgPheIle 577
         ::: |||||:::|||
Db 2572 ---ACGGAGCAACGACGAGCCATGGGCCCTGAGGTGAACGAGCCACTCAAGAGCAAGCTCAGA 2628

Qy   578 LeuProSerLeuAspAspProThrMetProAlaLeuAsnThrAlaPheSerThrGlnGlyLysArg 597
         :::||| :::|||
Db 2629 GGGCCCCCC--------------CCACAGGA-AACTTCGCAGTGATGCCAGGCGGAGG 2672

Qy   598 AlaArgAspGluTrpAspThrMetPheLysIle 608
         ||| :::|||
Db 2673 AAACGAGACCTCTTTCGTTCCTTCTAGAGAGGTC 2705

RESULT 7
US-08-742-923A-3
; Sequence 3, Application US/08742923A
; Patent No. 5869611
; GENERAL INFORMATION:
APPLICANT: Liu, Pu
APPLICANT: Collins, Francis S.
APPLICANT: Siciliano, Michael J.
APPLICANT: Claxton, David
TITLE OF INVENTION: Markers for Detection of Chromosome 16
TITLE OF INVENTION: Rearrangements
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: Harness, Dickey & Pierce, P.L.C.
STREET: P.O. Box 828
CITY: Bloomfield Hills
STATE: MI
COUNTRY: USA
ZIP: 48303
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/742,923A
FILING DATE: No. 5869611ember 1, 1996
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Smith, DeAnn F.
REGISTRATION NUMBER: 36683
REFERENCE/DOCKET NUMBER: 2115-00869DVC
TELECOMMUNICATION INFORMATION:
TELEPHONE: (810) 641-1600
TELEFAX: (810) 641-0270
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 2887 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear

```
;  MOLECULE TYPE:    cDNA to mRNA
;  HYPOTHETICAL:     NO
;  ANTI-SENSE:       NO
;  ORIGINAL SOURCE:
;     ORGANISM: Homo sapiens
;     INDIVIDUAL ISOLATE: Sample 1
;     TISSUE TYPE: Acute myelomonocytic leukemia, M4Eo
;     TISSUE TYPE: subtype (inv16)
;  POSITION IN GENOME:
;     CHROMOSOME/SEGMENT: 16[inv(16) (p13q22)]
;  FEATURE:
;     NAME/KEY: CDS
;     LOCATION: 1..2658
US-08-742-923A-3

Alignment Scores:
Pred. No.:              1.95e-06          Length:        2887
Score:                  159.00            Matches:       150
Percent Similarity:     38.30%            Conservative:  130
Best Local Similarity:  20.52%            Mismatches:    230
Query Match:            4.83%             Indels:        225
DB:                     2                 Gaps:          35

US-10-736-868-2 (1-643) x US-08-742-923A-3 (1-2887)
```

```
Qy   15 IleAlaProLeuSerAlaGlnSerProSerThrSerAspAlaProGlyAlaLeuLeuSer  34
           :::|||   |||||| |||||:::||
Db  769 GTGGGCTGCTCCCTGCAGTTCCCAGGACACCCCAGGAGTTG                     810

Qy   35 SerLeuValGlyLysSerHisGlnLysLeuProLeuAlaProSerMetGluAlaLeuGlu  54
            ::||||| |||||
Db  811 ---CTTCAAGAGAAGAAACCCGGCAGAGAGCTTCAACGTGCTCACGAAGCTGCCGCCAGCTGGAG 867

Qy   54                                                                54
Db  868 GAGGAGGCGGAACAGCTGCCAAGACCAGCTGACGACGAGGAGGATGGAGGCCAAGACCAGCAGAACCTG 927

Qy   55 ----------LeuMetGlyValGlnPheValAspAlaLeuAlaLeuIleLeuLysLysGlyGln 69
              :::        |||:::|||   |||||
Db  928 GAGCGGCCACCATCCTCCAACTCTCCACTCTCCAGCTCTCCGACTCCGAAG--AAGAAGTGCAG 984

Qy   70 --------------MetGluMetAlaLysValAlaGlyPheLysThrGlnLeu          81
              ::|||   |||
Db  985 GACTTTGCCAGCACCGTGGAAGAGGGGAAGAGAAGAGGTTCCAGAAGAGATC           1044

Qy   82 GluValLeuGluLysValHisProAspGln----------PheAspLysTyrLysLysLeu 98
           :::|||   :::        :::|||   :::
Db 1045 GAGGAGGCGGAACAGCTGCCAAGACCAGCTGACGACGAGGAGGATGGAGGCCAAGACCAGCAGAACC 1104

Qy   99 Lys-------------------ValAspAspAsnAlaPheIleAspAspValMetGlnAsnGlyAsn 132
           :::|||
Db 1105 AAGACAGGCGACGCCACCATCTTCCAAATACCGGATGAAGGGCTTATGATAAACTGGAAAAGACC 1164

Qy  113 GluMetAlaLysLeuGlnProLysSerGlyTyrAsnAlaPheIleAspAspLeuMetLysGlnAsnGly 132
           :::        ||||| |||
Db 1165 CTCGTGTCCAACCTGGAAAAGAACAGCAGAGAAGAGCAGAGG-----TTTGATCAGTTGTTAGCCGAG 1218

Qy  133 GlyIleProIleGlySer                                             139
           |||
Db 1219 GAGAAGAAACATCTTCTTCCAAATACGGCATGATGAGAGACAGCTGGAGGCAGAAGCCAGG 1278

Qy  140 ------------IleArgGlyLeuGluAspSerGluGlnIleIleAlaLysLysAspAlaLysLeuArgGlnIle 150
                       :::        ||||| |||::
Db 1279 GAGAAGGGAAACCAAGGCCCTGTCCCTGGCTGGGCCCCTTGAAGAGGCCCTTGGAAGCCAAA 1338

Qy  151 ArgAspMetGluAsnThrAspSerProSerGluGlnAsnIleAlaLysProAlaValMetAspAspLysLysPhe 170
           :::|||   |||
Db 1339 GAGGAACTGCAGCCGACCAAC-----AAAATGCTCAAAGCCCGAAATGGAA------   1383

Qy  171 GlnThrGlnIleLeuProGlyLeuValAlaAlaAsnMetIleAlaGlyLysLysAsnProPhe-- 189
           :::::::::::::::::::
Db 1384 GACCTGGTCAGTCAGGATGACGTGGGGC                                   1413
```

```
Qy  190 LysMetProGlnMetArgLysAlaGlnAlaAlaAlaAlaProSerSerValPheGlnGlnAla 209
           |||
Db 1414 AAGAAGTCCATGAGCTGGAAGTCCAAGCGGCCCTGGAGACCCCAGAGTGAGGAGATG       1473

Qy  210 LeuAlaGlnArgAlaMetLeuGlyLysAsnAlaProValAlaGlyLysAlaGlyArgGlyGluGlu 229
           :::
Db 1474 AAGACCAGCTGGAGGAGGCTGGAGACGACGAGCTGCAAGCCTCG--------GAGGAC       1521

Qy  230 GlnArgMetMetMetAsnArgValAspArgMetGln----------------Gln       243
           :::
Db 1522 GCCAAACTG----------CGGCTGGAAGTCAACATGCAGGCGCTCAAGGGCCAGTCGAA 1572

Qy  244 ArgGluLeuGlnGluGluAspAspAspAspLeuGluAspGluValProArg           263
           |||:::|||||||
Db 1573 AGGGATCTCCAAGCCCCGGGACGAGCAGAATGAGGAG--------AAG             1611

Qy  264 ArgArgSerSerAsp---------GlyGluProGlnSerGluAlaGluHisGlnArgArg   280
           |||||   :::::|||   |||
Db 1612 AGGAGGAGCAACTGCAGAGACAGCTTCACGAGTATGAGACGGAACTGGAGACGAGCGAAAC 1671

Qy  281 AspLeuAlaArgArgLeuLysSerSerProArgLeuLysGluLeuLeuGlnAlaAsnAlaGlu 300
           ::: |||
Db 1672 GAACGTGCCCTGGCAGCTGCAGCAAAGAAGAAGCTGAAGGGGGACCTGAAAGACCTGGAG 1731

Qy  301 ValGlnSerLeuLeuSerTyrGlnArgMetArgAspSerProLeuSerLysArgArgPro 320
           :::|||:::|||
Db 1732 CTTCAGGCC---GACTCTGCCATCAAGGGAGGAGGAGGAAGCCATCAAGCAGTTACGCAAA 1788

Qy  321 Leu------AlaMetAsnAsp---------GlyAspGluSerAlaPheArg---       332
           |||                |||
Db 1789 CTGCAGGCTCAGATGAATGAAGGACTTTCAAGAGAGTGCCCTGCCTCCAGAGAT         1848

Qy  333 AlaMetGluAlaArgAlaLysAspGlnLysSerGlnLysLeuValLeuGlyLeuHisGly  352
           |||   |||   |||
Db 1849 GAGATCTTTGCCACAGCCACAGCTGGAAGAGAATGAGAAGAAGAAAGCCAAGAGCTTG--- 1896

Qy  353 PheGlyGluSerAspAspAspAspLeuGluAspGluAspIleLeuAspProSerGlu     372
           :::||   |||||||
Db 1897 ------GAAGCAGACCTCATGCAGCTACAAGGAGAC----CTCGCCCGCTGAG         1941

Qy  373 AsnSerPheArgArgAlaProLeu----------ArgLeuSerGlyPheValGluValLys 389
           |||||
Db 1942 AGGGCTCGCAAACAAGCGGACTCCGAGAGAGGAACTGGCAGGAGGAGCTGGCCAGTAGC   2001

Qy  390 LeuLysSerAsnAspGluLeuLeuSerAlaAlaLeuAspProArgIleLysTyrArgValAspAsp 409
           |||
Db 2002 CTGTCGGGAAGGAACCACTCCAGGACCAGAAGCGCCGCCTGGAGCGCCGGATCGCCCAG     2061

Qy  410 ValGluLysTyrLeuAlaProLysProMetGluPheAsnProLysProGlnProGlyTyr 429
           :::||||| |||   :::||||||
Db 2062 CTGGAGGAGGAGCTGGAAGACGGAGGAGCAGGG-CAACATGGAGCGGCCATGAGCGACCGGGTCC-- 2119

Qy  430 PheAlaProArgLysIleProThr------------------------------Arg     438
           :::::|||
Db 2120 GCAAAGCCACCACAGCCAGGCCAGCAGCTGCCACAGAGCGCAGCACGG               2179

Qy  439 ProArgLysMet---LeuProLeuLeuIleGlyLeuTrpLysIleSerAspProLysProValGlnGluIle 457
           :::
Db 2180 CCCAGAAGAATGAGACGGTGCCC--------GGCAGCAGCTGCGAGCGGCAGAACAAGAGAGC 2230

Qy  458 ArgArgHisProSerThrGluPheLysIleAlaLysGluSerArgValLeuThrArgAsnLeu 477
           |||:::|||   :::
Db 2231 TCCGGCAGTCTCCACAATACCGAGATGGAGGGG-GCCGTCAAGTCC--------        2271

Qy  478 LysAsnAsnProSerLeuAlaAlaLeu----------PheMetAspAspLysLysLeu   492
                                      |||||
Db 2272 AAGTTCAAGTCCACCATCGCCGGCCTGGAGGCGCAAGATTGCACAGCTGGAGGAGCAGTC 2331

Qy  493 GluAsnThrLeuLysGlyArgGlnMetLeuThr----------Asp              503
           :::::|||   :::
Db 2332 GAGCAGGAGGCCAGAGAAACAGGCGACGCCACCAAGTCGCTCGAAGCAGAAAGACAAGAAG 2391

Qy  504 AspGlu                                                        505
Db
```

```
Qy   76  ------AlaPheLysThrGlnLeuGlu                                       82
                               ||||:::
Db  646  ATGCTAGAAAATTTACAGAACACTCAACACCCATGCTGCATTCCAGCCAAATTCCCAG        705

Qy   83  ValLeuGluLysValHisProAspGlnPheAspLysTyrLysLysLeuLysValAspAsp     102
            |||:||| ||||
Db  706  ATTGGTGAGGAAATGAGCCAGAACAGTTTCATAAAACAGTATCTGAAAAGCAGCAGGAG       765

Qy  103  LeuAlaAlaAspAlaValMetGlnGlnAla--GluMetAlaLysLeuGlnProLysSer      121
            :::   :::::|||    :::
Db  766  CTACTTAGGCAGCGCTGGAACGTGAAGCTGCGAGAACTTGAAGAAGGCTTCA              825

Qy  122  GlyAsnAlaPheIleAspMetLeuAsnGlyIle----------Pro                   135
            :::|||
Db  826  GCTGAGTCGGACGAGATGATCCATCCTGAGGGAGTGGCTTCCCGTGCCTCCTGAC          885

Qy  136  IleGlySerSerIleArgGlyLeuGluAspAlaIleLeuArgThrGlnArgAspMetGluAsn   155
            |||:::||||||
Db  886  TTTCAGAGCAGCTGGAGGACCAGAGCTGGAGCTGAGCAGACATTCGCCCAGAAAAAGC       945

Qy  156  ThrAspProSerGluGlnIleAlaLysAlaValMetAspLysPheGlnIleThrGlnIleIleLeu 175
            |||:::::
Db  946  TCCTCAATTTCTGAAGAG                                               963

Qy  176  ProGlyLeuValAlaAsnMetIleAlaGlyLysAsnProPheLysMetProGlnMet----    195
                       |||:::
Db  964  ------AAAGGTGACTCTGATGATGAGAAACCA                                990

Qy  196  ArgLysAlaGlnAlaAlaAlaProSerSerValPheGlnGlnAlaAlaLeuAlaGlnAlaAlaMet 215
            |||||| ||||:::
Db  991  AGGAAAGAGAAGACAGATCATCTAGGGTC--AGACAGGCAAGAGCAGCTAAACTGTCT      1047

Qy  216  LeuGlyLysLysAsnAlaProValAlaGlyValGlyArgGly------GlnArgMetMetGlu   233
            :::|||||||    |||:::
Db 1048  GAGGGC---AGCCACCTGCTGA-GGAGGAAGAGGATCAAGAAACACCTTCCAGAAACCT     1103

Qy  233  tMetAsnArgValAspGlnArgMet------GlnGlnArgGluLeuGlnLeuGluGlnAspGl   251
                                    |||:::
Db 1104  AAGGGTCAGCAGCAGATCGAAATTTGAAAACAGAGGAAGAGGAAGAGGAGGAGGA         1163

Qy  251  uAspAspAspAspLeuGluAspGluAspGluGlu------ValProArgArgArgSerSerAspGlyGluPr 271
            ||||:::::
Db 1164  GGAAGAAGATGATGAAGAAGAGAGAA----GGTGATGATGAGGG                    1202

Qy  271  oGlnSerGluLysGluHisGlnLysArgArgAspLeuAlaArgArgLeuArgArgLeuLysSerSerProAr 291
            |||   :::
Db 1203  ACAAAAATCTAGGGAG--------GCACCAAT                                1226

Qy  291  gLeuLysGlyLeuLeuGlnAlaAsnAlaGlyValGlnSerLeuLeuSerTyrGlnArgMetAr  311
            |||:::
Db 1227  CCTGAAAGAGTTTAAGGAGGAAGAGGGAA------GAGATACCTAGAGTAAA            1271

Qy  311  gAspSerProLeuSerLysArgArgProLeuAlaMetAsnAspGluArgGlu----        328
            |||    :::
Db 1272  ACCAGAGAGATGATGAGAGACCCAAAACAAGATCCCAGAACAGGAGGTGTTAGA          1331

Qy  329  ------SerAlaPheArgAlaMetGluAlaMetGluAlaLysAlaArgAlaLysLeuAspArgLysSerGlnLeu 346
                           ||||:::
Db 1332  GAGGAGGAGGAGATTTACAAGATTCCCAGGAAGAGGCTAGA-------AAAAGTCATCT   1382

Qy  346  uValLeuGlyLeuHisGlyTyrPheGlyGluSerAspAspGluAspGluLysGluAsp      366
            :::|||
Db 1383  GGGCC--------AGACAGCAGGAGGAGGAGAAATGAA                          1412

Qy  366  nLeuIleAspProSerGluAsnSerPheArgArgAlaProLeuArgLeuSerGlyPh        386
            |||:::
Db 1413  AACAACATCTCCCCTTGAGGAGGAAGAAGAAGAGAA-------ATAAAATCTTCACAAGGCTT 1466

Qy  386  eValGlyLeuLeuLysSer-----                                        399
            :::|||
Db 1467  AAAGGAAAAATCGAAGTCTTCCTTCCCTCCTGACTGACTGAAGAATGCTGAAGATGGAAAGAAGAGGCCTC 1526
```

```
Db 2392  CTGAAGCGAAATCTTGTCTGCAGGTGGAGGACGAGGCCAAGATGGCCGAGCAGTACAAGGAG  2451
                                                  ||||||
Qy  506  ------GlnLysGlyArgThrArgValLysThrIleArgAlaLeuProArgLeuPheGly     523
            :::||||||
Db 2452  CAGGCAGAGAAAGGCAATGCCAGGTCCAGCTCAAG-------AGGCAGCTGGAG          2502
Qy  524  AlaProThrAlaLysAlaGluMetIleAspAla-------LysValPheGlnAspIle       540
            :::::|||
Db 2503  GAGGCAGAGGAGGAGTCCAGCGCCATCAACGCCAACCGCCAGCAGAGCTGCAGCGGGAGCTG  2562
Qy  541  GluGluArgProIleProLeuProPhePheGluProLysGlyArgHisThrArgLeuArg     560
            :::|||
Db 2563  GATGCAGGCC--------------------                                  2571
Qy  561  TrpThrGlyAlaAsnGlu--------LysGluIleProGlyLeuProGlySerArgPheIle   577
            :::||||:::
Db 2572  ---ACGGAGCACAGCGGCCATGGCCCGTGACGGTGAACGACCAAGCTCAGA            2628
Qy  578  LeuProSerLeuAspProThrMetProAlaLeuAsnThrAlaPheSerThrGlnIleGlyArg  597
            ||||:::
Db 2629  GGGCCCCCC--------------CCACAGGA-AACTTCGCAGTGATGCACCAGGGCGAGG    2672
Qy  598  AlaArgAspGluTrpAspThrMetPheLysIle                               608
            |||   :::   :::
Db 2673  AAACGAGACACCTCTTTCGTTCCTTCCTTCTAGAGAGGTC                        2705
```

RESULT 8

US-09-949-016-1019

; Sequence 1019, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1019
; LENGTH: 4935
; TYPE: DNA
; ORGANISM: Human

US-09-949-016-1019

Alignment Scores:

Pred. No.:	
Score:	4.56e-06
Percent Similarity:	159.00
Best Local Similarity:	36.06%
Query Match:	22.39%
DB:	4.83%

Length:	4935
Matches:	159
Conservative:	97
Mismatches:	260
Indels:	196
Gaps:	35

US-10-736-868-2 (1-643) x US-09-949-016-1019 (1-4935)

```
Qy   23  ProSerThrSerAspAlaProGlyAlaProGlyAlaLeuLeuLeuSer---SerLeuValGlyLysSerHis 41
            ||||||||||
Db  484  CCGAGTACCTCACGCAAGATGGCGAGCTGGAGGAGGTGACTCTGGACGGGAAGCCTCTT     543
Qy   42  GlnLysGlyProLeuProLeuAlaProSerMetGluMetGlyLeuMetGlyValGlnPheVal   61
            |||||||||| :::
Db  544  CAGGGCGCTGCGGTGACCGACCTGAAGGCCGACTGGAGCAGCAGGC---------         591
Qy   62  AspAlaLeuIleLeuLysLysGlyGln---------MetGluMetAlaLysGly          75
            ||||:::
Db  592  ------CTTAGCCCAAGAGAGCGGGCCGCAAGAGAGTGCCCGTGGTCAAGCGGCTCAAAGGGGGCTCTA 645
```

```
;   PRIOR APPLICATION NUMBER: 60/237,768
;   PRIOR FILING DATE: 2000-10-03
;   PRIOR APPLICATION NUMBER: 60/231,498
;   PRIOR FILING DATE: 2000-09-08
;   NUMBER OF SEQ ID NOS: 207012
;   SOFTWARE: FastSEQ for Windows Version 4.0
;   SEQ ID NO 5054
;     LENGTH: 4935
;     TYPE: DNA
;     ORGANISM: Human
US-09-949-016-5054

Alignment Scores:
Pred. No.:                  4.56e-06      Length:           4935
Score:                      159.00        Matches:          159
Percent Similarity:         36.06%        Conservative:     97
Best Local Similarity:      22.39%        Mismatches:       260
Query Match:                4.83%         Indels:           196
                                          Gaps:             35

US-10-736-868-2 (1-643)  x  US-09-949-016-5054 (1-4935)

Qy    23  ProSerThrSerAspAlaProGlyAlaLeuLeuSer--------SerLeuValGlyLysSerHis  41
                :::|||||||||                          :::||| ||||   ||||||||
Db   485  CCGAGTACCTCACGCAAGATGGCGGAGCTGGAGGAGCTGACTCTGGACGGGAAGCCCTCTT  544

Qy    42  GlnLysLeuProLeuAlaProSerMetGluAlaLeuLeuLeuMetGlyValGlnPheVal  61
             ||||   |||| |||  |||  |||  ||| ||| |||  ||||||||| |||||
Db   545  CAGGCGCTGCGGGTGACCGACCTGAAGGCCGCACTGGAGCAGCGAGGC--------------  592

Qy    62  AspAlaLeuIleLeuLysGlyGln-----------MetGluMetAlaLysGly--------  75
             :::                                |||||||||||||||
Db   593  -------CTAGGCAAGAGCGGGCAGAAGAGTGCCCTGTCAAGCGGCTCAAGGGGCTA  646

Qy    76  ----------------------------------AlaPheLysThrGlnLeuGlu  82
                                              |||||||
Db   647  ATGCTAGAAAATTACAGAAACACTCAACACCCATGCTGCATTCCAGCCAAATTCCAG  706

Qy    83  ValLeuGluLysValHisProAspLysPheGlnPheLysLysLeuLysValAspAsp  102
             |||||                   :::         :::
Db   707  ATTGGTGAGGAAATGAGCCAGACAGTTTCATAAAACAGTATCTGAAAAGCAGCAGGAG  766

Qy   103  LeuAlaAlaAspAlaValMetGlnInAla---GluMetAlaLysLeuGlnProLysSer  121
             |||                      :::    ||| ||||||
Db   767  CTACTTAGGCAGCGTCGAAGCTGAAGCTCGAGAGCTGCAGAACTGAAGAAGCTTCA  826

Qy   122  GlyAsnAlaPheIleAspMetLeuAsnGlyAsnGlyIle------------------Pro  135
                                                                       |||
Db   827  GCTGAGTCGGGAGGACGAGATGATCATCCTGAGGGAGGTGGCTTCCCTGCCTCCTGAC  886

Qy   136  IleGlySerIleArgGlyLeuGluAspAlaIleArgThrGlnArgGlyMetGluAsn  155
             :::                 ||||      |||   |||
Db   887  TTTCAGGCAGCCTGGAGAGACCAGAGCTGGAGCTCAGCAGCATTCGCCCAGAAAAGC  946

Qy   156  ThrAspProSerGluGlnIleAlaLysAlaLysAlaValMetAspLysPheGlnThrGlnIleLeu  175
             ||||||||:::
Db   947  TCCTCAATTCTGAAGAG-------------------------------------------  964

Qy   176  ProGlyLeuValAlaAsnMetIleAlaGlyLysAsnProPheLysMetProGlnGlnMet  195
                                                                :::  :::
Db   965  ---------------------------------------AAAGGTGACTCTGATGATGAGAAAACCA  991

Qy   196  ArgLysAlaGlnAlaAlaProSerValPheGlnGlnInAlaLeuAlaGlnArgAlaMet  215
             ||||                                  |||
Db   992  AGGAAAGGAGAGAAAGACGATCATCTAGGGTC---AGACAGGCAAGAGCAGCTAAACTGTCT  1048

Qy   216  LeuGlyLysAsnAlaProValAlaGlyValGlyArgGly-GluGlu------GlnArgMetMet  233
             :::                |||   :::                |||
Db  1049  GAGGGC--AGCCAACCTGCTGA-GGAGGAAGGAGGATCAAGAAACACCTTCAGAAACCT  1104

Qy   233  tMetAsnArgValAspGlnArgMet----------GlnGlnArgGluLeuGlnGluGluAspGl  251
             :::          :::  :::             |||||||||:::|||
Db  1105  AGGGGTCAGAGCAGATCGAAATTTGAAAACAGAGGAGGAGGAGGAGGAGGAGGA  1164
```

```
Qy   399  aLeuAspArgIleLysTyrArgValAspAspValGluLysTyrLeuLeuAlaProLysProMe  419
             :||||                          :::||||
Db  1527  ACTTGTAGCCGTGCCAGACAAACTGCCAGCGAGGAGGAG------ACTCCTCCACCTTT  1580

Qy   419  t--------GluPheAsnProLysProGlnProGlyTyrPheAlaProLysIl  435
             :            :::       |||   |||       :::|||
Db  1581  ACTAACAAAGAAGCATCTTCTCCACCACTCCATCCCATAGCCAGGAAGAGAAAT  1640

Qy   435  e------------ProThrArgPro---------  439
             :::          |||
Db  1641  AGAGCCCATGGAAGGCCAGCCCCCCTGTCCTCATTCAGTTATCCTCCTAATACAGA  1700

Qy   440  ----------ArgLysMetLeu-----  443
                       |||||:::::||||
Db  1701  TGCTGACACCAGGGAGCTATTAGTATCTCAGCATACTGTCCAGTTGTGGAGGGCCTGTC  1760

Qy   444  -ProLeuIleGlySerAsp-ProLys--------ValGlnGluIleArgArgH  460
             |||||| ||||||  ||||||            :::  ||
Db  1761  TCCTTTGTCAAGTCCTTCAGACACCACCCAAGCACCAAAGCAGGAAAGTGCCAGGAGGA  1820

Qy   460  isProSerThrGluTrp--------------LysIleAlaLysGluSerArgValLeuT  475
             |:::                          :::  |||   :::
Db  1821  GAGTGTCCTGCCTCTGTTCAGAAAAGCACACTGGCTGACTACTCAGCCCAGAAGGATCT  1880

Qy   475  hrAsnLeuLysAsn-----------AsnProSerLeuAlaAlaLeu---------P  487
             :::                       |||||||||||:::
Db  1881  TGAACCTGAGTCAGACAGATCTGCTCAGCCCCTCTAAAAATTAGGAATTAGCACT  1940

Qy   487  heMetAspAspLysLeuGluAsnThrLeuGlyArgGlnMetLeuThrAspGlyGlnL  507
             :::  |||                 :::                  :::::::::::
Db  1941  GGCCAAAGGAATCACTGAAGAATGT-CTGAAA------CAGCCAATCTTTGGAACACAGAAGG  1993

Qy   507  ysGlyThrArgValLysThrIleArgAlaLeuProArgLeuPheGlyAlaProThrA  527
             :::       |||||                 |||  |||
Db  1994  AAGGCAGAGAGAGCTTCTCATCACCTTCTCCCCAAGCCACAGATTGAAACAGTCAGCTGATT  2053

Qy   527  laLysAlaGluMetIleAspProAlaLysValPheGlnAspIleGluGluArgProIleProP  547
             |||:::                    |||
Db  2054  CATCCCTCTAGCGCCGGTCCTCCTCCATCTTCCTCCTCAGTTCTAGATCAAGATCTCGCTCTC  2113

Qy   547  roLeuPhePheGluProLysGly---------ArgHisThrArgLeuArg--------  560
             |||                        |||||:::
Db  2114  CT------------GACAGTTCAGGTTCCGGTCTCATTCACCCGTCCAGATCCAAGCAGAGAG  2164

Qy   561  --------TrpThrGlyAlaAsnGluLysGluIleProGlyLeuGluSerArgP  576
                     |||:::            ||||
Db  2165  ATGTAGCCCAGGCACGTACTCATGCCAACCCGTGTAGACCCAAGATGGGGTCCAGAT  2224

Qy   576  heIleLeuProSerLeuAspProThrMetProAlaLeuAsnThrAlaPheSer------  593
             |||||:::                    :::
Db  2225  CAACATCAGAGTCCAGATCACAGGTCACAGTTCACGTTCTCGTTCAGCATCAAGCAACAGCA  2284

Qy   594  -----------ThrGlnGlyAlaArgAlaArgAspGluTrpAspThrMetPheLysIleProA  610
                         :::  |||||||||
Db  2285  GAAAATCTCTGAGCCCTGGAGTCTCCAGGGACAGC---AGCACCAGTATACTGAAACCA  2341

Qy   610  snAsnTrpAsnProGlyAspGluVal  618
             |||  ||||   ||||||
Db  2342  AAGATCCCTCTTCTGGTCAGGAGGTT  2367
```

RESULT 9
US-09-949-016-5054
;   Sequence 5054, Application US/09949016
;   Patent No. 6812339
; GENERAL INFORMATION:
;   APPLICANT: VENTER, J. Craig et al.
;   TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;   TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;   FILE REFERENCE: CL001307
;   CURRENT APPLICATION NUMBER: US/09/949,016
;   CURRENT FILING DATE: 2000-04-14
;   PRIOR APPLICATION NUMBER: 60/241,755
;   PRIOR FILING DATE: 2000-10-20

```
Qy  547 roLeuPheGluProLysGly-----ArgHisThrArgLeuArg-------------  560
        ::   |||                  |||||
Db 2115 CT--------------------GACAGTTCAGGTTCGTCCTTCATTCACCGTCAGATCCAAGCAGAGAG 2165

Qy  561 --------------TrpThrGlyAlaAsnGluLysGluIleProGlyLeuGlySerArgP 576
                      |||    ||||||       |||||
Db 2166 ATGTAGCCCAGGCACGTACTCATGCCAACCCTCGGTGTAGACCCAAGATGGCTCCAGAT 2225

Qy  576 heIleLeuProSerLeuAspProThrMetProAlaLeuAsnThrAlaPheSer-------  593
        |||                             ::: |||
Db 2226 CAACATCAGATCCAGATCAAGGTCACGTTCACGTTCTCGTTCAGCATCAAGCAACAGCA 2285

Qy  594 -------ThrGlnGlyAlaArgAspGluTrpAspThrMetPheLysIleProA  610
               :::  ||   |:||||||||       |||
Db 2286 GAAAATCTCGAGCCCTGGAGTCTCCAGGACAGC---AGCACCAGCTATACTGAAACCA 2342

Qy  610 snAsnTrpAsnProGlyAspGluVal  618
        :::  ||||| ||||||
Db 2343 AAGATCCCTTCTTCTGGTCAGGAGGTT 2368
```

RESULT 10
US-09-917-254-40
; Sequence 40, Application US/09917254
; Patent No. 6703204
; GENERAL INFORMATION:
; APPLICANT: Mutter, George
; APPLICANT: Baak, Jan
; TITLE OF INVENTION: Prognostic Classification of Breast Cancer
; FILE REFERENCE: B0801/7224(JRV)
; CURRENT APPLICATION NUMBER: US/09/917,254
; CURRENT FILING DATE: 2001-07-27
; PRIOR APPLICATION NUMBER: US 60/222,093
; PRIOR FILING DATE: 2000-07-28
; NUMBER OF SEQ ID NOS: 102
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 40
;   LENGTH: 5574
;   TYPE: DNA
;   ORGANISM: Homo Sapiens
US-09-917-254-40

Alignment Scores:
Pred. NO.:   Length: 5574
Score: 5.53e-06   Matches: 137
Percent Similarity: 159.00   Conservative: 121
Best Local Similarity: 39.15%   Mismatches: 212
Query Match: 20.79%   Indels: 192
DB: 4.83%   Gaps: 32
                     4

```
US-10-736-868-2 (1-643) x US-09-917-254-40 (1-5574)

Qy   15 IleAlaProLeuSerAlaGlnSerProSerThrSerAspAlaProGlyAlaLeuLeuSer  34
         ::||||||||  |||||||:::
Db 3583 GTGGCGTCCCTCAGTTCCCAGCTCCCAGGACACCCAGGAGCTG------------------ 3624

Qy   35 SerLeuValGlyLysSerHisGlnLysLeuProLeuAlaProSerMetGluAlaLeuGlu  54
         -------CTTCAGAAGAAACCCGGCAGAAGCTCAACGTGTCCGAAGCTGGCCAGCTGGAG 3681
Db 3625

Qy   54 --------------------------------------------------------------  54
Db   54

Qy   54 MetGluMetAlaLysGlyAlaPheLysThrGlnAsp  81
         GAGGAGCGGAACAGCCTGCAAGACCAGTGGACGAGGAGGAGATGAGGCCAAGCAGAACCTG 3741
Db 3682

Qy   55 -------LeuMetGlyValGlnPheValAspAlaArgLeuIleLeuLysLysGlyGln  69
Db 3742 CAGCCCATCTTTGGAACAGGAAGG----------------------------------- 3798

Qy   70 SerLeuValGlyThrArgValLysThrGlyGlyAlaLeuProArgMetGluMetAla  54
Db 3799 GACTTTGCCAGCACGTGGAAGCTTCGGAAGAGGGAGAAGAGGTTCCAGAAGAGAGATC 3858

Qy   82 GluValLeuGluLysValHisProAspGln----------PheAspLysTyrLysLysLeu  98
         ||| |||                                 :::||||  ||
```

```
Db   3859 GAGAACCTCACCCAGCAGTACGAGGAGAAGGCGGCCTTATGATAAACTGGAAAAGACC 3918
             |||
Qy     99 ...Lys--------------ValAspAspLeuAlaAlaAspAlaValMetGlnGlnAla 112
                              :::||||||||:||||  |||
Db   3919 AAGAACAGGCTTCAGCAGGAGCTGGACGACCTGGTTGTTGATTGGACAACCAGCCGCAA 3978

Qy    113 GluMetAlaLysLeuGlnProLysSerGlyLysAsnAlaPheIleAspMetLeuAsnGlyAsn 132
             :::::||||||||:||||::  |||
Db   3979 CTCGTGTCCAACCTGGAAAAGACAGCAGAGGAAA-------TTTGATCAGTTGTTAGCCGAG 4032

Qy    133 GlyIleProIleGlySerSer------ 139
             ||| |||
Db   4033 GAGAAAAACATCTCTTCCAAATACGGCGATGAGAGGGACAGCTGAGCAGAAGCCAGG 4092

Qy    140 ------------IleArgGlyLeuGluAspAlaIleArgThrGln 150
                          :::  |||:||:::|  |||
Db   4093 GAGAAGGAAACCAAGGCCCTGCCCTGGCTCGGGCCCTTGAAGAGCCAAA 4152

Qy    151 ArgAspMetGluAsnThrArgAspProSerGluGlnIleAlaLysAlaValMetAspLysPhe 170
             ||||::|||::::||||:::  |||
Db   4153 GAGGACTCGAGCGGCCAAC----AAAATGCTCAAAGCCGAAATGGAA------ 4197

Qy    171 GlnThrGlnIleIleLeuProGlyLeuValAlaAlaAsnMetIleAlaGlyLysAsnProPhe 189
                            :::::::|  |||||
Db   4198 ------------GACCTTGGTCAGCTCCAAGGATGACGTGGGC 4227

Qy    190 LysMetProGlnMetArgLysAlaAlaGlnAlaAlaGlnAlaProSerValPheGlnGlnAla 209
                 |||:  :::  |||
Db   4228 AAGAACGTCCATGAGCTGGAAGAGTCCAAGCGGCCCTGGAGACCCAGATGGCCCAGAGAGATG 4287

Qy    210 LeuAlaGlnAlaMetLeuGlyLysAsnAlaProValAlaGlyLyArgGlyArgGlyGlu 229
             |||                                       |||
Db   4288 AAGACCAGCTGGAAGAGCTGGAAGAGCTGCAAGCCACG---------GAGGAC 4335

Qy    230 GlnArgMetMetAsnArgValAspGlnArgMetGln----------Gln 243
             :::::::  |||
Db   4336 GCCAAACTG---CGGCTGGAAGTCAACATGCAAGGCGCTCAAGGGCCAGTTCGAA 4386

Qy    244 ArgGluLeuGlnGluGlnLysGluAspLysGluAspAspAspAspProArgGluGluAspGluValProArg 263
             ||||||||  |||
Db   4387 AGGGATCTCCAAGCCCGGGACGACCAGAATGAG-------GAGAAGAGGAGGCAACTGCAG 4440

Qy    264 ArgArgSerSerAspGlyProGlnInSerGlyLysAlaGluHisGlnHisArgArgAspLeuAla 283
             ::::||||||                  :::::::
Db   4441 AGACAGCTCACGAGTATGAGACAGGAACGGCGAAAGCAACGTGCCCTGGCA 4500

Qy    284 ArgArgLeuLeuSerSerProArgGlyLeuLysSerGlyLeuLeuGlnAsnAlaGluValGlnSer 303
             :::::||||::  ||||:::::
Db   4501 GCTCAGCAAGAAG-------AAGCTGGAAGGGGACCTGGAAGACCTTGGGAGCTTCAGGCC 4554

Qy    304 LeuLeuSerTyrGlnArgMetArgAspSerProLeuSerLysArgArgLeu----------- 321
             |||                            |||||
Db   4555 ---GACTCTGCCATCAAGGGGGAGGAGAAGCCATCAAGCAGTCTACGCCAAACTGCAGGCT 4611

Qy    322 AlaMetAsnAsp---------------------GluAspGluSerAlaPheArg---AlaMetGlu 335
             |||                               :::
Db   4612 CAGATGAAGAGACTTTCAAAGAGCTGGAAGATGCCGAAGATGAGATCTTT 4671

Qy    336 AlaArgAlaLysLeuAspProGlnLysSerGlnLeuValLeuHisGlyPheGlyGlu 355
             ||| :::|||||  |||
Db   4672 GCCACAGCCAAGAATGAACAGAGAAAAGCCAAGAGCTTG----------GAA 4713

Qy    356 SerAspAspAspGluGluAspGluGlnAspGluAsnLeuIleLeuAspProSerGluAsnSerPhe 375
             :::::|||||  |||||::
Db   4714 GCAGACCTCATGCAGCTACAAGAGGAC---------CTCGCCGCCGCTGGAGGGGCTCGC 4764

Qy    376 ArgArgAlaProLeu-------ArgLeuSerGlyPheValGluLysLysLeuLysLysSer 392
             ::::::||| |||                    |||
Db   4765 AAACAGACCGGAACCTCGAGAAGGAGGAGGAACTGGCAGAGGAGACTGGCCAGTTGTCGGGGA 4824

Qy    393 AsnAspGluLeuLysSerAlaLeuAsnAspArgIleLysTyrArgValAspAspValGluLys 412
             :::   |||||||:::||:|
Db   4825 AGGAACGCACTCCAGGACGAGAAGCGCGCCCTGGACGCCCGGATCGCCGCCCAGCTGGAGGAG 4884
```

```
Qy    413 TyrLeuAlaProLysProMetGluPheAsnProLysProGlnProGlyTyrPheAlaPro 432
             :::  ||||||
Db   4885 GAGCTGGAGGAGGAGGACAGGAGG--CAACATGAGGCCATGAGCGACCGGGTCC--GCAAAGCCA 4942

Qy    433 ArgLysIleProThr----------------------------------ArgProArgLys 441
             ||:::  |||
Db   4943 CACAGCAGCCGCCAGAGCTCAGCACGAGCTGGCCACAGAGCCGCAGCGACGCCCAGAAGA 5002

Qy    442 Met---LeuProLeuLeuIleGlySerAspProLysValGlnGluIleIleArgArgHis 460
             |||  :::                        :::
Db   5003 ATGAGAGTGCCC--------GGCAGCAGCTGCAGTCGAGCGGCAGAACAAGGAGCTCCGGACA 5053

Qy    461 ProSerThrGlyTrpLysIleAlaLysGluSerArgValLeuThrAsnLeuLysAsnAsn 480
             :::|||||||                                            :::||||
Db   5054 AGCTCCACGAGATGGAGGGGG--GCCGTCAAGTCC------------------AAGTTCAAG 5094

Qy    481 ProSerLeuAlaAlaLeu-------------------PheMetAspLysLeuGluAsnThr 495
             |||:::||||||||                      :::::::|||:||:|
Db   5095 TCCACCATGCCGGCGTTGGAGGCCAAGATTGCACAGCTGGAGGAGTCGAGCAGCAGGAG 5154

Qy    496 LeuLysGlyArgGlnMetLeuThr---------------------------------- 503
             ::: :::|||||
Db   5155 GCCAGAGAGAAACAGGCGGCCACCAGCTCGCTGAAGCAGAAGACAGAAGAGTCGAAGGAA 5214

Qy    504 --------------AspGlu----------------------------------Gln 506
                            |||||                                  :::
Db   5215 ATCTTGCTGCAGGTGGAGGACAGGCCCAGATGGCCCAGAGTACAGGAGGCAGGCAGCAGAG 5274

Qy    507 LysGlyArgThrArgValLysThrIleArgAlaLeuProArgLeuPheGlyAlaProThr 526
             |||
Db   5275 AAAGGCAATGCCAGGGTCAAGCAGCTCAAG--------------------AGGCAGCTGGAGGAGGCAGCAGAG 5325

Qy    527 AlaLysAlaGluMetIleAspAla---------------------LysValPheGlnIleAspIleGluGlu 542
             ||||:::::::|||  :::::::  ||||||||
Db   5326 GAGGAGTCCCAGCCGCATCAACCGCCAACCGCCAGGAAGTCGCAGCGGGAGCTGCAGCTGCGAGCTGGATGAG 5382
```

RESULT 11
US-09-949-016-1240
; Sequence 1240, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR APPLICATION DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1240
; LENGTH: 6861
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-1240

Alignment Scores:
Pred. No.:   Length: 6861
Score: 7.68e-06   Matches: 137
Percent Similarity: 159.00   Conservative: 121
Best Local Similarity: 39.15%   Mismatches: 212
Query Match: 20.79%   Indels: 192
DB: 4.83%   Gaps: 32
                              4

US-10-736-868-2 (1-643) x US-09-949-016-1240 (1-6861)

Qy   15 IleAlaProLeuSerAlaGlnSerProSerThrSerAspAlaProGlyAlaAlaLeuLeuSer 34
             ::: |||||:::::::::
Db   4825 AGGAACGCACTCCAGGACGAGAAGCGCGCCCTGGACGCCCGGATCGCCGCCCAGCTGGAGGAG 4884

```
Db          |||        |||||||||::|||                    |||  :::
4016  GTGGCGTCCCTCAGTTCCCAGCTCCAGGACACCCAGGAGCTG------------------    4057

Qy   35  SerLeuValGlyLysSerHisGlnLysLeuProLeuAlaProSerMetGluAlaLeuGlu    54
         |||  :::::: |||||||||  :::   |||
Db  4058  ---CTTCAAGAGAAACCCGGCCAGAAGCTTCAACGTGTCTACGAGAGCTGCGCCAGCTGGAG    4114

Qy   54  ------                                                           54
         |||
Db  4115  GAGGAGGCGGAACAGCCTGCGAAGACCAGCTGACGAGGACGATGAGGCCAAGCAGAACCTG    4174

Qy   55  ---------LeuMetGlyValGlnPheValAspAlaLeuIleLysLysGlyGln          69
                   :::   |||::|||     |||::::
Db  4175  GAGGGCCACCATCTCCACTCCAGCTCTCCGACTCGAAG--AAGAGCTGCAG            4231

Qy   70        MetGluMetAlaLeuAlaPheGlyThrGlnLeu                         81
              ||||||  ||| :::  :::::
Db  4232  GACTTGCCAGCACCGTGAAGCTTCGAGAGAGGGGAAGAAGAGGTTCCAGAAGGAGATC      4291

Qy   82  GluValLeuGluLysGln----------PheAspLysTyrLysLeu                  98
         ||| ::: ::  |||                :::
Db  4292  GAGAACCTCACCCAGCACGACGAGAAGGCGCCGCCTTATGATAAACTGGAAAAGACC       4351

Qy   99  Lys---------ValAspAspLeuLeuAlaAlaAspAlaValMetGlnGlnAla          112
         |||           :::|||||||:::||||||
Db  4352  AAGAACAGCAGCTTCAGCAGCGTGGACGAGCCTGGTTGTTGATTGGACAACCAGCGGCAA    4411

Qy  113  GluMetAlaLysLeuGlnProLysGlySerGlyAsnAlaPheIleAspMetLeuAsnGlyAsn   132
         ::::::  |||  ::::   |||
Db  4412  CTCGTGTCCAACCTGGAAAAGAAGCAGAGGAAA------TTTGATCAGTTGTTAGCCGAG     4465

Qy  133  GlyIleProIleGlySer-----                                         139
         |||
Db  4466  GAGAAAAACATCTCTTCCAAATACGCGGATGAGAGGACAGAGCTGAGGCAGAAGCCAGG     4525

Qy  140  -------IleArgGlyLeuGlnAspAlaIleArgThrGln                        150
                |||||| :::||||||:::::||
Db  4526  GAGAAGGGAAACCAAGGCCCTGCCCTGCCTCGGCGCCCTTGAAGCGGCCTTGGAAGCCAAA   4585

Qy  151  ArgAspMetGluAsnThrAspProSerGluGlnIleAlaLysAlaValMetAspLysPhe    170
         ::::::|||     |||   :::::
Db  4586  GAGGAACTCGAGGCGGACCAAC-------AAAATGCTCAAAGCCGAAATGGAA--------    4630

Qy  171  GlnThrGlnIleLeuProGlyLeuValAlaAlaAsnMetIleAlaGlyLysAsnProPhe--   189
                                      :::::::
Db  4631  --------GACCTGGTCAGCTCCAAGGATGACGTGGGC                         4660

Qy  190  LysMetProGlnIleMetProGlnMetArgValAlaAspValMetGln-             209
         ||| :::||||||||::::::    :::  :::
Db  4661  AAGAACGTTCCATGAGCTGGAGAGTCCAAGCCGCCCTGGAGACCAGATGGAGGAGATG     4720

Qy  210  LeuAlaGlnArgAlaMetLeuGlyLysAsnAlaProValAlaGlyAlaGlyArgGlyGlu    229
         :::   |||       |||:::|||
Db  4721  AAGACGCAGCTGGAAGACGCTGGAGGACGACGAGCTGCAAGCCACG-------GAGGAC    4768

Qy  230  GlnArgMetMetAsnArgValAspValMetGln-----------Gln                243
            :::
Db  4769  GCCAAACTG--------CGGCTGGAAGTCAACATGCAGCGCGCTCAAGGGCCACAGTTCGAA   4819

Qy  244  ArgGluLeuGlnGluGluAspGluAspAspAspLeuGluAspGluValProArg         263
         |||||||||||||  :::::              |||
Db  4820  AGGGATCTTCCAAGCGCCGGAGGAGCAGAGAATGAG-----GAGAAGAGGAGCAACTGCAG   4873

Qy  264  ArgArgSerSerAspGlyGluProArgLeuLysGlnAlaGluHisGlnArgArgAspLeuAla    283
         ||::::   ::: |||  |||       :::::|||  :::
Db  4874  AGACAGCTTCACCAGTATGAGACGACGGAAGACCTGGAACAGCGAAAGCAACGTGCCCTGCCA   4933

Qy  284  ArgArgLeuLysSerSerProArgLeuLysLeuGlnLeuLeuGlnAlaGluValGlnSer    303
         |||  |||:::||||||      ::::||  :::::::||:::::|||
Db  4934  GCTGCAGCCAGCAGAGAG-----AAGCTGGAAGGGGGACCTGAAAGACCTCGAGCTTCAGGCC    4987

Qy  304  LeuLeuSerTyrGlnArgMetArgAspSerProLeuSerArgArgProLeu------       321
         :::   :::         |||
Db
```

```
Db   4988  ---GACTCTGCCATCAAGGGAGGAGGAGGAGAAGCCATCAAGCAGCTACGCAGCAAACTGCAGGCT    5044

Qy   322  AlaMetAsnAsp----------GluAspGluSerAlaPheArg---AlaMetGlu         335
          |||||||||||                |||||||
Db   5045  CAGATGAAGGACTTTCAAAGAGACCTTGGAAGATGCCCGTGCCTCCAGAGATGAGATCTTT    5104

Qy   336  AlaArgAlaLysLeuAspGlnLysSerGlnLysLeuGlyLeuValLeuGlyLeuHisGlyPheGlyGlu    355
          |||  |||||| :::||||||||||||::|||
Db   5105  GCCCAGCCCAAAGAACATGAGAGAAGAAAGCCAAGACCTTG-----------GAA    5146

Qy   356  SerAspAspAspProMetGluAspGluAspAlaAsnLeuIleLysAspProSerGluSerPhe    375
          :::::::::          ||||||    :::  |||||
Db   5147  GCGAGACCTCATGCAGCTACAGACAGAC---------CTCCGCGCCGCCTGAGAGGGCTCGC    5197

Qy   376  ArgArgAlaProLeu----------ArgLeuSerSerGlyPheValGluValGluLeuLysSer    392
          ::::|||   :::       |||||
Db   5198  AAACAAGCCGGACCTTCGAGAAGGAGGAACTGGCAGAGGACTGGCCAGTAGCCTGTCGGGA    5257

Qy   393  AsnAspGluLeuLysSerAlaLeuAspArgIleLysTyrArgValLysAspAspValGluLys    412
          :::  |||::::|||                        :::||:::||
Db   5258  AGGAGACGCACTCCAGGACGAGAAGGCCCTGGAGGCCCGATCGCCCAGCTGGAGGAG    5317

Qy   413  TyrLeuAlaProLysProProMetGluPheAsnProLysProGlnProGlyTyrPheAlaPro    432
          ::||||||                           |||
Db   5318  GAGCTGGAGGAGGAGCAGG--CAACATGGAGGCCATGAGCGACCGGGTCC--GCAAAGCCA    5375

Qy   433  ArgLysIleProThr----------ArgProArgLys                          441
          |||||||           |||||||||
Db   5376  CACAGCAGGCCGAGCAGCTCAGCAACGAGCTGGCCACAGCGCAGCAGCGCCCAGAGA    5435

Qy   442  Met---LeuProLeuLeuIleGlySerAspAspProLysProLysGlnIleGluGlnIleLeuArgArgHis    460
          :::                         ||||||::::::|||
Db   5436  ATGAGAGTGCCC-------------GGCAGCAGCGCCAGAACAACAAGGAGCTCCGAGCA    5486

Qy   461  ProSerThrGluTrpLysIleIleAlaLysValLeuSerArgValLeuThrAsnLeuLysAsnAsn    480
          |||||                     ||||||:::
Db   5487  AGCTCCACGAGATGGAGGG-GCCGTCAGTCC---------------AAGTTCAAG    5527

Qy   481  ProSerLeuAlaAlaLeu----------PheMetAspAspLysLeuGluAsnThr         495
                                     |||||||||||
Db   5528  TCCACCATGCGCGCGCTGAGGCCCAAGATTGCACAGTCGAGGAGCAGGTCGAGCAGGAG    5587

Qy   496  LeuLysGlyArgGlnMetLeuThr-                                       503
          ::: :::|||||||
Db   5588  GCCAGAGAAACAGCGGCGCCACCAAGTCGCTGAAGCAGCAGAAAGAAGAAGCTGAAGGAA   5647

Qy   504  --------AspGlu-------------------------------Gln               506
                                                        :::
Db   5648  ATCTTGCTCCAGGTGGAGGACGACGCGCAAGCAGCGCCGAGCAGTACAAGGAGCAGGAGAG    5707

Qy   507  LysGlyArgThrArgValTyrThrIleArgAlaLeuProArgLeuPheGlyAlaProThr    526
          ||||||:::::::||||||:::     |||
Db   5708  AAAGGCAATGCCAGGGTCAAGCAGCTCAAG----------AGGAGCTGAGGAGGAGCAGAG    5758

Qy   527  AlaLysAlaGluMetIleAspAla--------LysValPheGlnAspIleGluGlu        542
          ::::::|||           |||::::|||:::||
Db   5759  GAGGAGTCCCAGGTCCCAGCCATCAACCGCAACCCGCAGAAGCTGCAGCCCGGGGAGCTGGATGAG   5815
```

RESULT 12
US-09-949-016-1241
; Sequence 1241, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03

```
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1241
; LENGTH: 6861
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-1241

Alignment Scores:
Pred. No.:
Score:                 159.00          Length:        6861
Percent Similarity:    39.15%          Matches:       137
Best Local Similarity: 20.79%          Conservative:  121
Query Match:                           Mismatches:    212
DB:                                    Indels:        192
              7.68e-06                 Gaps:          32
```

```
US-10-736-868-2 (1-643) x US-09-949-016-1241 (1-6861)

Qy    15  IleAlaProLeuSerAlaGlnSerProSerThrSerAspAlaProGlyAlaLeuLeuSer   34
              ::||:||||    :||||:|||
Db  4016  GTGGCGTCCCTCAGTTCCAGCTCCCAGGACACCCAGGAGCTG------------------ 4057

Qy    35  SerLeuValGlyLysSerHisGlnLysLeuProLeuAlaProSerMetGluAlaLeuGlu   54
              :::::||    ||||||||||    ::|:::    |||||
Db  4058  ---CTTCAAGAAGAAACCCGGCAGAAGCTCAACGTGTCTACGAAGCTGCCGCAGCTGGAG 4114

Qy    54                                                                 54
Db  4115  GAGGAGGCGGAACAGCCTGCAGCAGCCAGTGGAGGAGGATGAGAGCCAAGCAGAGACCTG 4174

Qy    55  ------LeuMetGlyValGlnPheValAspAlaLeuIleLeuLysLysGlyGlyGln   69
                  :|||    |||    ::|||||||
Db  4175  GAGGCGCCACATCTCCAACATCCAGCTCTCCGACTCGAAG------AAGAAGTGCAG   4231

Qy    70      ------MetGluMetAlaLysGlyAlaPheLysThrGlnLeu   81
              :|:    ||||||    ::||    :::||:::
Db  4232  GACTTTGCCAGCACCGGAAGCTCTGAAGAGGGGAAGAGAGAGGTTCCAGAAGGAGATC   4291

Qy    82  GluValGluGluLysValHisProAspGln------------PheAspLysTyrLysLysLeu   98
              |||    :||||    ::|    :::|||
Db  4292  GAGAACCTCACCAGCATACGAGGAGAAGGGCGCCGTTATGATAAACTGGAAAAGACC   4351

Qy    99  Lys------------ValAspAspLeuAlaAlaAspAlaValMetGlnAlaValAla   112
              |||    :::||||||||||    :|||    :|:
Db  4352  AAGAACAGGCTTCAGCAGGAGCTGGACGAGACCTGGTTGTGTTGGACAACCAGCGGCAA   4411

Qy   113  GluMetAlaLysGlnLeuProLysSerGlyAlaAsnAlaPheIleAspMetLeuAsnGlyAsn   132
              :::::||||||||    ||||||    |||
Db  4412  CTCGTGTCCAACCTGGAAAAGAACAGCAGAGGAA------TTTGATCAGTTGTTAGCCGAG   4465

Qy   133  GlyIleProIleGlySerSer   139
              |||||    |||
Db  4466  GAGAAAAACATCTCTTCCAAATACGCGGATGAGAGGACGCTGAGGCAGCAGAAGCCAGG   4525

Qy   140  ------------IleArgGlyLeuGluAspAlaIleAlaArgThrGln   150
                       ::|||:::||    ::||:::
Db  4526  GAGAAGGAAACCAAGGGCCCGTCCCTGGCTCCGGAGCTCCTTGAAGAGCCCAAA   4585

Qy   151  ArgAspMetGluAsnThrAspProSerGluGlnIleAlaLysAlaValMetAspLysPhe   170
              |||||:::::|    :::||    |||||
Db  4586  GAGGAACTCGAGCGGGACCAAC------AAAATGCTCAAAGCCGAAATGGAA   4630

Qy   171  GlnThrGlnIleLeuProGlyLeuValAlaAlaAsnMetIleAlaGlyLysAsnProPhe---   189
                    :::::|    |||||
Db  4631  ------------------GACCTGGTCAGCTCCCAAGGATGACGTGGGC   4660

Qy   190  LysMetProGlnGlnMetArgLysArgIleAsnAlaGlnAlaAlaProSerSerValPheGlnGlnAla   209
              |||||:::::|    |||||:::    :::::
Db  4661  AAGAACGTCCATGAGCTGGAGAAGCTGGAGCTCCAGGACACCGAGATGGAGGAGATG   4720

Qy   210  LeuAlaGlnArgAlaMetLeuGlyLysAsnAlaProValAlaGlyProValAlaGlyArgGlyArgGluGlu   229
              ||||||||
```

```
Db  4721  AAGACGCCAGCTGGAGGAGCTGGAGGACGAGCTGCAAGCCACG------GAGGAC   4768

Qy   230  GlnArgMetMetAsnArgValAspGlnArgMetGln   243
              ||||||    :::    ::|
Db  4769  GCCAAACTG------------CGGCTGGAAGTCAACATGCAGGGCCTCAGGGCCGTCGAA   4819

Qy   244  ArgGluLeuGlnGluGluAspGluAspAspLeuAspGluGluAspValProArg   263
              |||:::||:||||||||    ::::|||
Db  4820  AGGGATCTCCAAGCCCGGACGACGCAGAATGAG------GAGAAGAGGAGGAACTGCAG   4873

Qy   264  ArgArgSerSerAspGlyGluProGlnSerGlyGluAlaGluHisGlnArgArgAspLeuAla   283
              ||||||    :::::::||    :::::::||||||
Db  4874  AGACAGCTTCACGAGTATGAGACGAAGACTGGAAGACGCGAGAAAGCGAAAGCACGTGCCCTGGCA   4933

Qy   284  ArgArgLeuLysSerSerProArgLeuLysGluLeuGlnAsnAlaGluValGlnSer   303
              :::||||||||:::::|    |||||    ||||||
Db  4934  GCTGCAGCAAAGAAG------AAGCTGGAAGAGGGGAACTGGAAGACCTGGAGCTTCAGGCC   4987

Qy   304  LeuLeuSerTyrGlnArgMetArgAspSerProLeuSerLysArgArgProLeu---   321
              :||    :::::||:::
Db  4988  ---GACTCTGCCATCAAGGGGAGGGAGGAGGGAAGCCATCAACGACTACGACCAAACTCAGCGCT   5044

Qy   322  AlaMetAsnAsp------------GluAspGluSerAlaPheArg---AlaMetGlu   335
              ||||||:::::||    |||||||||||
Db  5045  CAGATGAAGGACTTTCAAAGAGAGCTGAAGATGCCCGTGCCTCCGAGAGATGAGATCTTT   5104

Qy   336  AlaArgAlaLysLeuLys------------GlnLysSerGlnLeuValLeuGlyLysLeuHisGlyPheGlyGlu   355
              ||||||:::||||||    ||||||||||||
Db  5105  GCCACAGCCAAAGAAGACCTCCAAAGAGAATGAGAGAGAAGCCAAGAGCTTG------GAA   5146

Qy   356  SerAspAspAspGluGluGluAspGluAspGluAsnLeuIleAspProSerGluAsnSerPhe   375
              :::::|||    |||:::::|    :::||||||
Db  5147  GCAGACCTCATGAGCTACAAGGAC------CTCGCGCCGCCTGAGAGGGGCTCGC   5197

Qy   376  ArgArgAlaProLeu------------ArgLeuSerGlyPheValGluLysLeuLeuLysSer   392
              ||||||||    |||:::|||||
Db  5198  AACAAGCCGACCTCCAGGAGGAGGAACTGGCAGAGATTCTGGAGAGCTGGAGGGA   5257

Qy   393  AsnAspGluLeuLysSerAlaLeuAspArgArgIleLeuTyrArgValAspGluLys   412
              :::||||||||:::||||||
Db  5258  AGGAAGGACCCACTCCAGGACGAAGCGCGCCTGGAGGCCCGATCGCCAGCTGGAGGAG   5317

Qy   413  TyrLeuAlaProLysProMetGluPheAsnProLysProGlnProGlyTyrProAlaPro   432
              |||||    :::::|||||    ::|||
Db  5318  GAGCTGGAGGAGGCCAGGG---CAACATGGAGGCCATCAGCGACCCGGGTCC--CGAAAGCCA   5375

Qy   433  ArgLysIleProThr---   441
              :::    |||:::
Db  5376  CACAGAGGCCGAGCAGCTCAGCAACGAGCTGGCCACAGAGCGCAGCAGCGCCCAGAAGA   5435

Qy   442  Met---LeuProLeuLeuIleGlySerAspProLysValGlnAsnGluIleArgArgHis   460
              |||    :::
Db  5436  ATGAGAGTGCCC------GGCCAGCAGTCGAGCTGGAGCCCGAGAACAAGGACGTCCGAGGCCA   5486

Qy   461  ProSerThrGluTrpLysTrpIleAlaLysGluSerArgValLeuThrAsnLeuLysAsnAsn   480
              |||:::::|||||||    :::|||||
Db  5487  AGCTCCACGAGATGGAGGG---GCCGTCAAGTCC-------AAGTTCAAG   5527

Qy   481  ProSerLeuAlaAlaLeu------------PheMetAspLysLeuGluAsnThr   495
              |||:::::||    :::::::::
Db  5528  TCCACCATCGCGGCGCTGGAGGCCCAAGATTGCACAGCTGGAGGAGGTCGAGCAGGAG   5587

Qy   496  LeuLysGlyArgGlnMetLeuThr   503
              ::::|||
Db  5588  GCCCAAGACAAACAGGCGGCCACCAAGTCGCTGAAGCAGACAAGAAGTGAAGGAA   5647

Qy   504  ------AspGlu   506
              |||
Db  5648  ATCTTGTCAGTGAGGACGTGGAGGCCGAAGATGGCCAAGGACCAGCAGCAGAGG   5707

Qy   507  LysGlyArgThrArgValLysThrIleArgAlaLeuArgAlaLeuProArgLeuPheGlyAlaProThr   526
              ||||||:::||||
```

Db   5708 AAAGGCAATGCCAGGGTCAAGCAGCTCAAG--------AGGCAGCTGGAGGAGGCAGAG 5758

Qy   527 AlaLysAlaGluMetIleAspAla-------LysValPheGluAspIleGluGlu 542
           :::::::: |||::|| |||:: ::::::::::::::|||

Db   5759 GAGGAGTCCCAGCCCATCAACGCCACCAGGAGCTGCAGCCGGAGCTGGATGAG 5815

RESULT 13
US-09-949-016-1242
; Sequence 1242, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT APPLICATION FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR APPLICATION FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR APPLICATION FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR APPLICATION FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SEQ ID NO 1242
; LENGTH: 6861
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-1242

Alignment Scores:
Pred. No.:   7.68e-06   Length: 6861
Score: 159.00   Matches: 137
Percent Similarity: 39.15%   Conservative: 121
Best Local Similarity: 20.79%   Mismatches: 212
Query Match: 4.83%   Indels: 192
DB: 4   Gaps: 32

US-10-736-868-2 (1-643) x US-09-949-016-1242 (1-6861)

Qy   15 IleAlaProLeuSerAlaGlnSerProSerThrSerAspAlaProGlyAlaLeuLeuSer 34
           ::||| ||||||||||::||||
Db   4016 GTGGCGGTCCCTCAGTTCCCAGCTCCAGGACACCCAGGAGCTG----------------- 4057

Qy   35 SerLeuValGlyLysSerHisGlnLysProLeuProLeuAlaProSerMetGluAlaLeuGlu 54
           ||| |||||| :::||| ::: |||||||
Db   4058 ---CTTCAAGAGAAACCCGGCGAAGCTCAACGTGTCTACGAAGCTGCGCCAGCTGGAG 4114

Qy   54   54
           :::
Db   4115 GAGGAGCGGAACAGCCTGCAAGACCAGCTGGACGAGGAGGATGGAGGCCAAGCAGAACCTG 4174

Qy   55 --------LeuMetGlyValGlnPheValAspAlaLeuIleLeuLysGlyGln 69
           ::: ||:|| ||||| ::: ||||||
Db   4175 GAGCGCCACCATCTCCACTCCACATCCAGCTCTCCAGCTCCGAAG--AAGAAGCTGCAG 4231

Qy   70 ------------MetGluMetAlaLysGlyAlaLysPheLysThrGlnLeu 81
           :::||| ||| :::::::
Db   4232 GACTTTGCCAGCACCGTGCAGAGCTCTGGAAGCTCTGAAGAGGGGAAGGAAGAAGGAGATC 4291

Qy   82 GluValLeuGluLysGlyValHisProAspGln--------PheAspLysTyrLysLysLeu 98
           |||||||:::::   ::: |||||:::
Db   4292 GAGAACGCCAACCCCAGCAGTACGAGGACGAGGAGGAGGCCGCCTTATGCATGAGGAAAGACC 4351

Qy   99 Lys------------ValAspAspAlaLeuAlaLeuAlaAlaAlaValMetGlnGlnAla 112
           ||| :::|||||||||||| |||::
Db   4352 AAGAAACAGCCTTCACCCAGCTACGAGGAGCTGGACTTGGTTCTTGTTGGACAACCAGCCGCAA 4411

Qy   113 GluMetAlaLysLeuGlnProLysSerArgGlyAsnAlaPheIleAspMetLeuAlaAsnGlyAsn 132
           :::||| |||:::: |||||| |||
Db   4412 CTCCGTGTCCAACCTGAAAAGAGCAGAGGAAA------TTTGATCAGTTGTTAGCCGAG 4465

```
Db  12710 GAA------GAGGCGTACGACCTGCCCGGCCTCGATACACCGCTCAAGGCCCTCCTT 12657
Qy    102 AspLeuAlaAlaAspAlaArgValMetGlnGlnAlaGluMetAlaAlaLysGln-------- 118
            :::           ||||||
Db  12656 GAGACGGTTGGAGTCGCCGCCCTGAGCAACGAACAGTCCCTCTTGAAGGCAATGCAG 12597
Qy    119 ------------ProLysSerGlyAsnAlaPheIleAspMetLeuAsn 130
                              ||||||:||||
Db  12596 GGGGCCTACCGCCAGTTCTTCCAGCACGGGAAAG------CTCGTCACC 12549
Qy    131 GlyAsnGlyIlePro---IleGlySerSerIleArgGlyLeu---GluAspAlaIleArg 148
            ||:::  |||||     :::::|||     |||||||||||
Db  12548 GGAAGCGACCTCGATCGTGAGGGTGCTCGCATGCAGGAGGTCGCGGAGGCAGCTCGG 12489
Qy    149 ---ThrGlnArgAspMetGluAsnThrAspProSerGluGlnIleAlaLysAlaAlaValMet 167
               ||||||
Db  12488 TTGACTCAGCGGTTGGCGGAG------ATCGAGCAGCTTGCCAGCGAGAGCATG 12441
Qy    168 AspLysPheGlnInThrGlnIleLeuProGlyLeuValAlaAsnMetIleAlaGlyLysLysAsn 187
            |||           |||:::
Db  12440 GACCTCCGGCCAGAAGCAGCTC------AGGCAGCTGGAGGAGGCAGAG 12420
Qy    188 ProPheLysMetProGlnMetArgLysAlaGlnAlaAlaAlaProSerValPheGln 207
            :::   :::|||:::
Db  12419 ------TCCCTGGCGGAGAACGAGCACCTATGAGCAGCGGCTCGCTAGCTTG---CAG 12369
Qy    208 GlnAlaAlaLeuAlaGlnAlaArgAlaMetLeuGlyLysAsnAlaProValAlaGlyIleGlyArgGly 227
            :::    :::||||:::     |||
Db  12368 GAGCGGCCATCGCCGGAGGAGAGCTGGAGCGGCTCCGGGAAGAAGGACGTC 12309
Qy    228 GluGluGlnArgMetMetMetAsnArgValAspGlnArg--------Met 241
            |||:::||        :::|||
Db  12308 GAGGCGTCAGCGGAAAAGAGTGCCAGGCCGGACCAGCAGCAGCAGAGCTTCGCGCTC 12249
Qy    242 GlnGlnArgGluGlnGluGlnGluAsp------GluAspAspAspAspLeuGluAsp 258
            ||:::||        :::||||||
Db  12248 CAGCAGAAGGTGCTCCAACGGCTCTGGCTCGCAGCAGCCGGCCTTGCAGCAG 12189
Qy    259 GluAspValProArgArgArgSerSerAspGluProGlnSerGlnAlaGlnGluHisGln 278
               |||
Db  12188 GCCCAGGGACTCGTCCGGCCGCCGAAGACAGGGGTCTCGCGAAGCACGGAGGCGCACCAAG 12129
Qy    279 ArgArgAspLeuArgArgLeuLysSerSerProArgLeuLysGluLeuGlnAsn 298
               |||
Db  12128 ------GCTCAGCAG 12120
Qy    299 AlaGluValGlnSerLeuLeuSerTyrGlnArgMetArgAspSerProLeuSerLysArg 318
            :::     |||:::|||
Db  12119 GGGGAGGTGGAG------CTGCGGGAGCAGCGCGTGGAACGAGCAGCGACGACTCCTCGAA--- 12069
Qy    319 ArgProLeuAlaMetAsnAspGluAspGluSerAlaPheArgAlaMetGluAlaArgAla 338
            |||:::||||||
Db  12068 ------------GATGCGGCGCGCT 12057
Qy    339 LysLeuAspLysSerGlnLeuLeuValLeuGlyLeuLeuHisGlyPheGlyGluSerAspAsp 358
            :::|||||:::|||
Db  12056 TCCCTGGCGGAGCAGCGGAAGTTG------ 12033
Qy    359 AspGluAspGluGlnUAspGluAsnGluIleAspProSerGluAsnSerPheArgArgAla 378
            :::     :::
Db  12032 GACGGGCTCGTCAAGCAGGGGACCGCGGGTGGAGACCCGGAGAATGCCACGCAG 11973
Qy    379 ProLeuArgLeuSerGlyPheValGluLysLeuLysSerAsnAspGlu------ 395
               |||||:::||||||:::
Db  11972 ACTGCGGCGCTGAAGCGCCTTCCAGACGCAGAGGTAAGCGCGGAGGTGCAGCGC 11913
Qy    396 ------LeuLysSerAlaLeuAspArgIleLeuTyrArgValAspAspValGluLysTyrLeu 414
               :::||||||
Db  11912 AAGCTGAATCAGCCCCTGACACAGCTGGACTGTGTCGAGGGGGTCGAGGTC---TCCTTCACT 11856
Qy    415 AlaProLysProMetGluPheAsnProLysProGlyTyrPheAlaProArgLys 434
               |||                :::||:::
Db  11855 GCCGAGAGCCCCGCAGCCTATCGAGTGGGAGGCCCAG---GGTCACCACCACCGCACAAG 11799
```

```
Db   5436 ATGAGAGTGCCC------------GGCAGCAGCTCGAGCGGCAGAACAAGAGGAGCTCCGGAGCA 5486
Qy    461 ProSerThrGluTrpLysIleAlaAlaLysGluSerArgValLeuThrArgAsnLeuLysAsnAsn 480
            :::|||:::              |||
Db   5487 AGCTTCCACGAGATGGGAGGGGG-GCCGTCAAGTCC------AAGGTTCAAG 5527
Qy    481 ProSerLeuAlaAlaLeu------PheMetAspLysLeuGluAsnThr 495
            :::|||||        |||:::::||||||:::
Db   5528 TCCACCATGCCGGCGGCCCAGAGGGGCCGGCCAAGATTGCACAGCAGTTGCAGGAGCAGCAGGAG 5587
Qy    496 LeuLysGlyArgGlnMetLeuThr--------- 503
            |||:::|||:::
Db   5588 GCCAGAGAGAAACAGCGCGGCGCCACCAAGTCGCTGAAGCAGAAAGACAAGAGCTGAAGGAA 5647
Qy    504 ------AspGlu--------Gln 506
            :::   |||     |||
Db   5648 ATCTTCGTCAGGTGAGGAGGCCCACGAGCGGCCAGATGGCCAGTACAAGGAGCACAGGAGCCAGAG 5707
Qy    507 LysGlyArgThrArgValLysThrIleArgAlaLeuProArgLeuPheGlyAlaProThr 526
            |||||||
Db   5708 AAAGGCAATGCCACGAGGTCAAGCAGCTCAAG------AGGCAGCTGGAGGAGGCAGAG 5758
Qy    527 AlaLysAlaGluMetIleAspAla------LysValPheGlnAspIleGluGlu 542
            ||||||:::|||     ||:::|||
Db   5759 GAGGAGTCCCAGCGCATCAACGCCAACGCCCAAGACGCCAGGAAGCTGCAGCGGGGGAGCTGGATGAG 5815
```

```
Alignment Scores:
Pred. No.:              4.46e-05        Length:        19269
Score:                  158.50          Matches:       125
Percent Similarity:     37.32%          Conservative:  106
Best Local Similarity:  20.19%          Mismatches:    215
Query Match:            4.82%           Indels:        173
                                        Gaps:          29

US-10-736-868-2 (1-643) x US-09-902-540-1175 (1-19269)

Qy     27 AspAlaProGlyValAlaLeuLeuSerSerLeuValGlyLysSerHisGlnLysLeuProLeu 46
             :::|||||:::
Db  12884 GACGACGGCCCAGTAGCCTCCTCGATGAGTGGACACGGGGAGCTCGGTTCGAGCGCG------CTC 12831
Qy     47 AlaProSerMetGluAlaLeuGluLeuMetGluValGlnPheVal 61
             |||           :::|||
Db  12830 GCTGAATCGCAGGAGGCGGATGCGTTCGTCCTTCCTGTCTCCAGCGTGCCCGGC 12771
Qy     62 AspAlaLeuIleLeuLysGlyGlnMetGluAlaLysAlaPheLysThrGlnLeu 81
             :::                |||:::
Db  12770 AACGCGCTGACGACAAGGTGGGCAACTGGGGCTTGGCTTCGTCGGTTGCTCTGGTTGAATCAGTCG 12711
Qy     82 GluValLeuGluLysLysValHisProAspGlnPheAsnPheGlnAspLysTyrLysLysLeuLysLysValLysAsp 101
             |||:::
Db  11855
```

Percent Similarity: 38.31%   Conservative: 127
Best Local Similarity: 19.16%   Mismatches: 211
Query Match: 4.77%   Indels: 198
DB: Gaps: 32

US-10-736-868-2 (1-643) x US-08-875-435B-1 (1-6175)

```
Qy    15 IleAlaProLeuSerAlaGlnSerProSerThrSerAspAlaProGlyAlaAlaLeuLeuSer   34
            ::||||    :    :  ::      :  :
Db  4032 GTGGCTTCCCTTGGATCCAGCTTCAGGACACCCAAGAGCTG-----------            4073

Qy    35 SerLeuValGlyLysSerHisGlnLysLeuProLeuAlaProSerMetGluAlaLeuGlu     54
            ||||||  :::::::   |||||||
Db  4074 ---CTCCAAGAGAAACCCGGCAGAGCTTCAATGTGTCTACCAAGCTGCGTCAGTTGGAA    4130

Qy    54                                                                  54
Db  4131 GATGAAAGGAACAGCCTGCGAGGACCCAGCTGGATGAGGAGGAGCTAAGCAAAACCTG     4190

Qy    55 -----------LeuMetGlyValGlnPheValAspAlaLeuLeuLysLysGlyGln         69
                        :: |||||
Db  4191 GAGGGCCATGTGTCTCAACACTGAACATTCAGCTCTCAAG---AAGAAGCTGCAG        4247

Qy    70                       MetGluMetAlaLysGlyAlaLysGlyAlaAlaPheLysThrGlnLeu   81
                               ||||||  :: :::
Db  4248 GACTTTGCAAGTACCATCGAGGTCATGAGGAGGAGGAAGAAGAGTTACAGAAAGAGATG    4307

Qy    82 GluValLeuGluLysValHisProAspGln------------PheAspLysTyrLysLysLeu  98
            |||||| :: :::
Db  4308 GAGGGCCCTGCAGCCAGCAGTACGAGGAGGCGCTGCCTATGACAAACTGGAGAAAACC     4367

Qy    99 Lys----------ValAspAspLeuAlaAlaAspLeuValMetGlnAlaValAlaMetGlnAla  112
            |||
Db  4368 AAGAACAGGCTCCAGCAGGAGCTCGATGAGTCGTGGTCACCAGCGGCAA              4427

Qy   113 GluMetAlaLysLeuGlnProLysSerGlyAsnAlaPheIleAspMetLeuAsnGlyAsn   132
            :::::
Db  4428 CTGTATCCAATCTGAAAAGAAGCAGAAGAAA-----TTTGACCAGTTGTTAGCTGAG      4481

Qy   133 GlyIleProIleGlySerSer-----                                      139
            ||| |||
Db  4482 GAGAAGAACATCTCCTCCAAGTATGCGAGTGAGAGAGCCAGG                     4541

Qy   140 ------------------------------IleArgGlyLeuGluAspAlaIleAlaArgThrGln  150
                                          |||:::
Db  4542 GAAAAGGAGACAAAGGCTTTGTCTTCTAGCCCGGGCCCCTGGAGGAAGCCAAA          4601

Qy   151 ArgAspMetGluAsnThrAspProSerGluGlnIleGluAlaLysAlaAlaValMetAspLysPhe  170
            ::::::|||
Db  4602 GAAGAGCTGGAGAGGACCAAC------AAGATGTCAAAGTCAGGATGGAA------         4646

Qy   171 GlnThrGlnIleLeuProGlyLeuValAlaAsnMet---IleAlaGlyLysAsnProPhe   189
            |||:::|||
Db  4647 ------GACCTGGTCAGCTCCAGGATGATGTAGGCAGAACGTGCAT                  4688

Qy   190 LysMetProGlnMetArg------                                        200
            :::::||| :: |||
Db  4689 GAACTGGAGAAGTCCAAGCGTGCCTTGGACCCAGATGGAAGAGATGAAAACCCAGCTG      4748

Qy   201 AlaProSerSerValPheGlnAlaLeuLeuAlaLeuAlaArgAlaMetLeuGlyLysAsnAla   220
            ::: |||
Db  4749 GAGGAGTCGGAGGATGACGTGCAGGCC--ACTGAGGATGCCAAGTCCGGCTAGAGGTC     4805

Qy   221 ProValAlaGlyLeuArgGlyGluGlnGlnArgMetMetMetAsnArgValAspGlnArg   240
            ::: :::|||
Db  4806 AACATGCAGGCCCTCAAGGGCCAGTTGAACGCGATCTCCAGGCTCCGGCGGATGAACAGAT   4865

Qy   241 MetGln---GlnArgGluLeuGlnGluLeuGlu---AspGluArgLeuGluLeuAspGluAsp  258
            ::: :::
Db  4866 GAGGAGAAGAGGAGGCAGCTACAGCGGCAGCTACAGCGAGTATGACAGACAGACTGAAGAT   4925

Qy   259 GluAspValProArgArgArgSerArgSerAspGlyProGlnSerGluGluHisGln      278
            :::::|||  ::: :::|||
```

```
Qy   435 IleProThrArgProArgLysMetLeuProLeuLeu----------                446
            ::||  ||:: :::
Db 11798 CTCACAAGGAGGAGCAGAAGGTCTTCTCGGGGTGCTGGTGCTGGTGCACTGCCATCAAG  11739

Qy   447 ----------------------IleGlySerAspProLysVal                    453
                                   ||||||
Db 11738 GGAGTGGGTGAAGTCCCGCTGCGCACTGCGGCGCCGAGGAGATTGGC------AAGCTC  11688

Qy   454 GlnGluGluIleArgArgHisProSer---------                            462
            :::  ||:::: ::|||
Db 11687 GAAGCGGGAGGTCGGAGAAGGACCTGCGCGCGCCTCCATGACGACGGAGTC           11628

Qy   463 ---ThrGluTrpLysIleAlaLysGluSerArgValLeuThr                      475
            ||||||  ::: ||||||
Db 11627 GATGATGTGGTGGGTCTGCGCGCGGAGTGG------GCCGTCCCAGCAGGTCGTGCTTCAG  11574

Qy   476 AsnLeuLysAsnAsnProSerLeuAlaAlaLeuPheMetAspAspLysGluGluAsnThr   495
            ||| :::::|||
Db 11573 GAGCTGTCGAAGTACGAGGAGGTACGACGGAGCCCAGTTCTTGGAGGAC------AACAAA  11523

Qy   496 LeuLysGlyArgGlnMetLeuThrArgAspGluGlnLysGlyArgThrArgValLysThrIle  515
            :::::::||||| ::: :: |||
Db 11522 GTCGAGAGGGTCGAGGTGCTCAAGGAACAGCGTGAGCAGGCACGG---AAGACG--       11469

Qy   516 ArgAlaLeuProArgLeuPheGlyAlaProThrAlaLysAlaGluMetIleAspAlaLys   535
            |||||| |||
Db 11468 -------GGGGCACTGAGCGCCCAGCTCAACCTTTCGCGG-----                 11436

Qy   536 ValPheGlnAspIleGluGluArgProIlePro--------ProLeuPhePheGluPro    552
            ||| |||
Db 11435 ------GAGGAGCTACAGACCACCCACCCCATTTCCCGAGCGCTGAAGGCGCTGACGGAGGAACTC  11382

Qy   553 LysGlyArgHisThrArgLeuArgTrpThrGlyAlaAsnGluLysGluIleProGly      571
            :: ||| :::||||
Db 11381 AAGGCGCTGTCAGCCTCCAGGAGTGAAGGCCCCGGACCAAGGCGCCGGAGGAGGTGGAAGGA  11325
```

Alignment Scores: 1.06e-05   Length: 6175
Pred. No.:   157.00   Matches: 127
score:

```
Db      4926  GAA------CGGAAGCAGAGAGCTCTGGCGCGCAGCTAAGAAGAAGAGCTGGAAGGGGAC  4979

Qy       279  ArgArgAspLeuAlaArgArgLeuLysSerProArgLeuLysGlu------LeuLeu     296
                  ::|||||               |||:::           ::|||
Db      4980  CTAAAAGACCTAGAGCTCCAGGCTGACTCAGCCATCAAAGGAGGGAGGAAGCCATCAAG  5039

Qy       297  GlnAsnAlaGluValGlnSer---LeuLeuSerTyrGlnArg---MetArgAspSerPro  314
                 |||   :::||||||||:::       :::|||||     :::   |||
Db      5040  CAGCTTCGAAAACTGCAGGCTGCAGATGAACGACTTCCAAAGAGACTGGATGATGCCCGT  5099

Qy       315  LeuSerLysArgArgProLeuAlaMetAsnAspGluSerAlaPheArgAlaMet        334
                 |||:::               |||     :::|||||:::        :::::::::
Db      5100  GCCTCCAGGGATGAGATCTTTGCCACCTCAAAGAGAATGAGAAGAAAGCCAAGAGTCTG  5159

Qy       335  GluAla------------Arg                                         337
                 |||||                :::
Db      5160  GAGGCAGACCTCATGCAGCTCCAAGAGACCTGGCCAGCAGCTGAGAGCTCGCAAGCAA  5219

Qy       338  AlaLysLeuAspGlnLys--------SerGlnLeuValLeuGlyLeuHisGlyPheGly  354
                 |||||            |||         :::|||
Db      5220  GCTGACCTGGAAGGAGGAGAGCTGCCGAGCAGCTGGCTAGCAGCTTGTCAGGAGGAAT  5279

Qy       355  GluSerAspAsp------------------------------AspGluAsp          361
                 |||                                       :::|||
Db      5280  ACACTGCAGGATGAGAAGCCGCCGCCTGGAGGCAAGGATCGCCCAACTAGAGGAGGAGCTG  5339

Qy       362  GluGluAspGluAsnLeuIleAspProSerGluAsnSerPheArgArgAlaProLeu--  380
                 |||||:::::::::||               :::
Db      5340  GAGGAAGACGAGGCCAACATGGAGGCCATGAGTAGTGATAGAGTTACGCAAGGCCACACTGCAG  5399

Qy       381  -----ArgLeuSerSerGlyPheValGluLys--------LeuLysSerAsnAspGlu  395
                      :::|||||||||               :::|||
Db      5400  GCTGAGCAACTGAGCAATGAGCTGGCCACAGAACGCAGCACGTCAGAAGAATGAGAGC  5459

Qy       396  LeuLysSerAlaLeuAspArg------------IleLysTyrArgValAspValGlu  411
                 :::        ||||:::|||          :::::::    :::|||||||
Db      5460  GCACGGCAACACGCTGAGCGCCAGAACAAGGAACTGCGAAGCAAGTTGCACGGAGGTAGAA  5519

Qy       412  LysTyrLeuAlaProLysProMetGluPheMetGluProLysAsnProLysProGlyPheTyrPheAla  431

Db      5519                                                                    5519

Qy       432  ProArgLysIleProThrArgProArgLysMetLeuProLeuLeuIleGlySerAspPro  451
                 :::                    |||         :::
Db      5520  ---GGTGCTGTCAAAGCCAAGCTCAAGTCCACTGTTGCGGCGCTGAGGCCAAGATTGCA  5576

Qy       452  LysValGlnGlnLysGlyGlnThrArgValLysThrProSerThrGluTrpGluThrPLysIleAlaLysGluSer  471
                 :::::::::|||::||:::    |||:::|||||    :::
Db      5577  CAGCTGGAGGAGAACAGCAGTTGAACAGGAG-----------GCCAGAGAGAAA  5615

Qy       472  ArgValLeuThr---AsnLeuLysAsnAsnProSerLeuAlaLeuAlaLeuPheMetAspAsp  490
                 :::|||          |||:::|||
Db      5616  CAGGCCGCCACCAAGTCGCTGAAGCAAAAG----------------GACAAG  5651

Qy       491  LysLeuGluAsnThrLeu----------------LysGlyArgGlnMetLeu  502
                 :::|||||||               |||
Db      5652  AAGCTAAAGAGAGGTCCTGCTCGCTGCCAGGTTGGAGGATGAGCGGCTGGAGGATGACGCAGAGCAGTACAAG  5711

Qy       503  ThrAspGluGlnLysGlyArgThrArgValLysThrIleArgAlaLeuProArgLeuPhe  522
                 :::|||||:::|||||       |||:::|||||       |||
Db      5712  GAGCAGGCAGAGAAAGAAACCCAAGGTCAAGCAGCAGTGAG------AGGCAGCTG  5762

Qy       523  GlyAlaProThrAlaLysAlaGluMetIleAspAla----------LysValPheGlnIleAsp  539
                 :::|||||||       ||||:::|||
Db      5763  GAAGAGGCAGAGGAGGAGTCCCAGTGCATCAACGCCCAACCCAGAGAAGCTGCAGCCGGGGAG  5822

Qy       540  IleGluGlu  542
                 :::::::|||
Db      5823  CTAGAATGAG  5831
```

Search completed: July 18, 2005, 16:15:46
Job time : 448 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: July 18, 2005, 12:25:31 ; Search time 2544 Seconds
(without alignments)
1603.522 Million cell updates/sec

Title: US-10-736-868-2
Perfect score: 3291
Sequence: 1 MILFLFLLLGFCIAPLSA.........SKTKRFVGGNGAFDMPALGL 643

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 717243 seqs, 3172129809 residues

Total number of hits satisfying chosen parameters: 14346486

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=x1h
-Q=/cgn2_1/USPTO_spool/US10736868/runat_14072005_105511_2334/app_query.fasta_1.839
-DB=Published Applications NA -QMT=fastap -SUFFIX=rnpb -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100
-THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=2000000000 -USER=US10736868 @CGN_1_1_480 @runat_14072005_105511_2334
-NCPU=6 -ICPU=3 -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100
-LONGLOG -DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications NA:*
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq:*
10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq:*
11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:*
12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
13: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq:*
14: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:*
15: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq:*
16: /cgn2_6/ptodata/2/pubpna/US10D_PUBCOMB.seq:*
17: /cgn2_6/ptodata/2/pubpna/US10E_PUBCOMB.seq:*
18: /cgn2_6/ptodata/2/pubpna/US10F_PUBCOMB.seq:*
19: /cgn2_6/ptodata/2/pubpna/US10G_PUBCOMB.seq:*
20: /cgn2_6/ptodata/2/pubpna/US10H_PUBCOMB.seq:*
21: /cgn2_6/ptodata/2/pubpna/US10I_PUBCOMB.seq:*
22: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
23: /cgn2_6/ptodata/2/pubpna/US11A_PUBCOMB.seq:*
24: /cgn2_6/ptodata/2/pubpna/US11_NEW_PUB.seq:*
25: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
26: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	3291	100.0	1932	21	US-10-736-868-1	Sequence 1, Appli
2	176.5	5.4	7396	17	US-10-028-248A-35	Sequence 35, Appl
3	176.5	5.4	7396	17	US-10-107-782-35	Sequence 35, Appl
4	173.5	5.3	7274	21	US-10-696-909A-48	Sequence 48, Appl
5	169.5	5.2	2791	20	US-10-425-115-63082	Sequence 63082, A
6	169.5	5.2	6126	17	US-10-188-248-25	Sequence 25, Appl
7	169.5	5.2	6151	18	US-10-302-172-528	Sequence 528, App
8	165	5.0	3603	17	US-10-369-493-27636	Sequence 27636, A
9	163.5	5.0	2932	19	US-10-747-065-1	Sequence 1, Appli
10	163.5	5.0	6354	15	US-10-084-817-158	Sequence 158, App
11	163	5.0	4041	17	US-10-264-049-543	Sequence 543, App
12	162.5	4.9	3257	17	US-10-369-493-30630	Sequence 30630, A
13	162.5	4.9	5883	21	US-10-956-157-859	Sequence 859, App
14	162	4.9	3320	19	US-10-283-975A-279	Sequence 279, App
15	159	4.8	3388	9	US-09-967-456-1602	Sequence 1602, Ap
16	159	4.8	3388	9	US-09-967-768A-245	Sequence 245, App
17	159	4.8	3388	9	US-09-954-531-988	Sequence 988, App
18	159	4.8	3388	10	US-09-873-367C-85	Sequence 1382, Ap
19	159	4.8	3388	18	US-10-240-425-1200	Sequence 85, Appl
20	159	4.8	3388	21	US-10-843-641A-85	Sequence 1200, Ap
21	159	4.8	3388	21	US-10-843-641A-2055	Sequence 85, Appl
22	159	4.8	3388	21	US-10-843-641A-2449	Sequence 2055, Ap
23	159	4.8	3388	21	US-10-843-641A-4629	Sequence 2449, Ap
24	159	4.8	3388	21	US-10-843-641A-6390	Sequence 4629, Ap
25	159	4.8	3388	21	US-10-887-553A-800	Sequence 6390, App
26	159	4.8	4935	21	US-10-956-157-1980	Sequence 800, App
27	159	4.8	4935	21	US-09-927-597-1	Sequence 1980, Ap
28	159	4.8	5835	10	US-10-486-057-1	Sequence 1, Appli
29	159	4.8	5835	22	US-09-927-597-3	Sequence 1, Appl
30	159	4.8	5937	10	US-10-486-057-3	Sequence 3, Appli
31	159	4.8	5937	22	US-10-171-311-161	Sequence 3, Appl
32	159	4.8	6861	14	US-10-341-434-102	Sequence 161, App
33	159	4.8	6861	17	US-10-923-035-18	Sequence 102, App
34	159	4.8	6861	22	US-10-171-311-163	Sequence 18, Appl
35	159	4.8	6900	14	US-10-764-425-13	Sequence 163, Appl
36	159	4.8	6900	22	US-10-116-802-14	Sequence 13, Appl
c 37	159	4.8	11065	14	US-10-116-802-14	Sequence 14, Appl
c 38	159	4.8	11065	14	US-10-437-963-1292	Sequence 14, Appl
39	158	4.8	2836	19	US-10-316-253-43	Sequence 1292, Ap
40	157.5	4.8	2348	16	US-10-282-122A-31788	Sequence 43, Appl
41	157.5	4.8	3486	17	US-10-171-581-352	Sequence 31788, A
42	157	4.8	6401	15	US-10-177-293-314	Sequence 352, App
43	157	4.8	6401	15	US-10-765-700-42	Sequence 314, App
44	157	4.8	6856	22	US-09-814-353-21993	Sequence 42, Appl
45	157	4.8	7677	10		Sequence 21993, A

ALIGNMENTS

RESULT 1
US-10-736-868-1
; Sequence 1, Application US/10736868
; Publication No. US20050079160A1
; GENERAL INFORMATION:
; APPLICANT: Solomon, Aaron
; APPLICANT: Morimoto, Richard
; APPLICANT: Beitel, Greg
; TITLE OF INVENTION: OSR-1 Nucleic Acids and Proteins
; FILE REFERENCE: NWESTERN-08451
; CURRENT APPLICATION NUMBER: US/10/736,868
; CURRENT FILING DATE: 2003-12-16
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1
; LENGTH: 1932

us-10-736-868-2.rnpb

```
;
; TYPE: DNA
; ORGANISM: Caenorhabditis elegans
US-10-736-868-1

Alignment Scores:
Pred. No.:
Score:                    1.41e-316    Length:              1932
                          3291.00      Matches:             643
Percent Similarity:       100.00%      Conservative:        0
Best Local Similarity:    100.00%      Mismatches:          0
Query Match:                           Indels:              0
DB:                                    Gaps:                0
                                                             21

US-10-736-868-2 (1-643)  x  US-10-736-868-1 (1-1932)

Qy    1 MetIleLeuPheLeuLeuPheLeuLeuLeuGlyPheCysIleAlaProLeuSerAla   20
Db    1 ATGATTTTATTTTTATTTTTATTCCTGTTGTTAGGATTTTGTATCCGCCACCATTATCGGCC   60

Qy   21 GlnSerProSerThrSerAspAlaProGlyAlaLeuAlaLeuLeuSerSerLeuValGlyLysSer   40
Db   61 CAAATCTCCATCGACTTCCGATGCTCCGGGAGCTTTGTTGTCATCTCTCGTAGGTAAAAGC   120

Qy   41 HisGlnLysLeuProLeuAlaProSerMetGluAlaLeuGluLeuMetGluValGlnPhe   60
Db  121 CATCAAAAACTACCACTGGCTCCCATCAATGGAAGCCTTGGAACTGATGGGTGTTCAATTT   180

Qy   61 ValAspAlaLeuIleLysLysGlyGlnMetGluMetAlaLysAlaPheLysThrGln   80
Db  181 GTTGATGCTCTCATCAAAAAAGGTCAAATGGAAATGGCAAAAGCATTTAAGACTCAA   240

Qy   81 LeuGluValLeuGluLysLysValHisProAspAspPheAspLysLysTyrLysLysLeuLysVal  100
Db  241 TTAGAAGTTCTAGAAAAGAAAGTACATCCTGATGATTTTGATAAGAAATACAAAAAGTT   300

Qy  101 AspAspLeuAlaLysAlaValMetGlnGlnAsnGlyGlyLeuGlnLysAlaLysGlnProLys  120
Db  301 GATGATTTGGCAGCTGCAGTCATGCAACAGAATGGCAACAGGAGGCCGAGATGCCGATGAGT   360

Qy  121 SerGlyAsnAlaPheIleAspMetLeuAsnGlyIleAsnGlyIleProIleGlySerSerIle  140
Db  361 TCAGGAAATGCATTTATTGATATGTTGAATGGAATCAATGGAATCCCAATTGGTAGCAGTATT   420

Qy  141 ArgGlyGluGluProLeuAspAlaIleGluArgThrGlnArgAspMetGluAsnThrArgAspProSerGlu  160
Db  421 CGTGGAGAATTAGAAGATGCTATCGAACGACAGAGAGATATGGAAAATACGCGATCCGTCCGAA   480

Qy  161 GlnIleAlaLysAlaValMetAspLysPheGlnIleLeuGlnIleLeuProGlyLeuValAla  180
Db  481 CAGATTGCCAAAGCCGTAATGGACAAATTTCAAACAAATTCTCCAGGACTCGTTGCA   540

Qy  181 AsnMetIleAlaGlyTyrAspAsnProPheLysMetProGlnThrGlnIleMetArgGlyAlaGlnAla  200
Db  541 AATATGATCGCTGGCAAGAGATGTTCCCATTGCTCAAAGAGAGCAATGTTAGGTAAAAATGCC   600

Qy  201 AlaProSerSerValPheGluGlnIleAlaLeuAlaGlnAlaMetLeuGlyLysAsnAla  220
Db  601 GCTCCATCGTCAGTGTTCGAGGAGATTGCTCTTGCTCAAGCTCTTGGTAAAAATGCC   660

Qy  221 ProValAlaGlyTyrArgGlyGluGluGlnMetMetAsnArgValAspGlnAsnArg  240
Db  661 CCAGTTGCCGGTGAAGAGCAGATGATGAATCGGATTGGAACCAGGTGACCAAAGA   720

Qy  241 MetGlnGlnAsnArgGlnGluGluGlyGluAspGluAspAspGluLeuGluAspGluAsp  260
Db  721 ATGCAACAAAATAGACAAGAAGAAGGAGAAGATGAAGATGATGAGGACGAGGAT   780

Qy  261 ValProValArgArgSerSerSerAspGlyGluProGlnSerGluLeuGluGlnLeuHisGlnArgArg  280
Db  781 GTACCCGTAAGGAGAAGAAGTTCCGATGGAGAACCACAAAGTGAAGCAGAGCAGCAATTCAGAGAAGA   840

Qy  281 AspLeuAlaLeuArgGluLysLysSerSerProArgLeuLysGluLeuGlnLeuGlnAsnAlaGlu  300
Db  841 GATTTAGCCAGGAGATTGAAAAGTAGTCCTAGATTAAAAGAGCTTTTACAGAATGCGGGAA   900

Qy  301 ValGlnSerLeuLeuSerTyrGlnArgMetArgAspSerProLeuSerLysArgArgPro  320
Db  901 GTTCAATCATTGCTCTTCTCTTACCAACGAATGAGGGATTCTCCACTGAGCAAGCCAAGGCCT   960

Qy  321 LeuAlaMetAsnAspGluSerAlaGluAlaPheArgAlaMetGluAlaArgAlaLysLeu  340
Db  961 TTGGCTATGAACGATGAGAGTGCAGAAGTGCATTCCGCGCCAATGGAGGCTCGTGCAAAACTA  1020

Qy  341 AspGlnLysSerGlnLeuValLeuGlyLeuHisGlyLyPheGlySerAspAspAspGlu  360
Db 1021 GATCAAAAATCTCAACTTGTCTTGGGTCTCCATGGTTTTGGAGAGTCTGATGATGAA  1080

Qy  361 AspGluGluAspGluAsnLeuIleAspProSerGluAsnSerPheArgArgAlaProLeu  380
Db 1081 GACGAAGAAGATGAAAATTTGATTGATCCATCTGAAAATTCATTCCGTGTCGCACCACTT  1140

Qy  381 ArgLeuSerGlyPheValGluValGlyLeuLysLeuLysSerAsnAspGluLysSerAlaLeu  400
Db 1141 CGTCTTTCTGGATTCGTTGAGAAAGTAAAGTCAAATGATGAAGTGCATTG  1200

Qy  401 AspArgIleLeuLysTyrArgValAlaAspValTyrArgLeuAlaProLysLysProMetGlu  420
Db 1201 GACAGAATTAAAATATCGAGTTGATGAGTGGAAAGTATCTTGCTCCAAAGCCGATGGAA  1260

Qy  421 PheAsnProLysProGlnTyrPheAlaProLysIleProThrArgLysIleLeuProThrArgArg  440
Db 1261 TTCAATCCAAAACCTCAGCCTGGCTACTTTGCTCCACTGAAAATCCCAACAAGAACCACGT  1320

Qy  441 LysMetLeuProLeuLeuIleGlyLeuGlnGluLeuIleLeuArgArgHis  460
Db 1321 AAAATGCTTCCATTATTAATTGGATCTGCAACAAGAGTTTGACAAATTGAAGAATAAT  1380

Qy  461 ProSerThrGluTrpLysIleGlyLySerGluArgValLeuThrArgAsnLeuLysAsnAsn  480
Db 1381 CCAAGTACCGAATGGAAAATTGGAAAATCGCAAGAGAGTGTCAAAACAATTCGTCATTACCAAGA  1440

Qy  481 ProSerLeuAlaAlaLeuPheMetAspAspLysSerLeuGluAsnThrLeuGlyGluThrLeuGlnArgGln  500
Db 1441 CCAAGTCTTGCTGCACTGTTCATGGATGATAAATCATTGGAAAATACACTTGAAGGAGGCAA  1500

Qy  501 MetLeuThrAspGluGlnLeuLysGlyTyrArgThrArgValLysThrArgIleIleArgAlaLeuProArg  520
Db 1501 ATGTTAACTGATGAACAGAAAGGTAGAACAAGATTCGTGCATTACAAGA  1560

Qy  521 LeuPheGlyAlaProThrAlaLysValAlaGlyMetIleAspAlaLeuAlaLysValPheGlnIle  540
Db 1561 CTGTTCGGTGCACCAACTGCAAAAGTTGAAAGTCAAAGTATTCCAAGATATT  1620

Qy  541 GluGluLeuArgProIleProProLeuPhePheGluLeuProGlyLeuArgGlyHisThrArgLeuArg  560
Db 1621 GAAGAACTGCGACCAATCCCCATTGTTTTTGAACCAGAAGGAGCATACGAGATTGAGA  1680

Qy  561 TrpThrGlyAlaAsnGluLysLeuGlyIleProGlyLeuGlySerArgPheIleGluProSer  580
Db 1681 TGGACTGGAGCAAATGAAAAGAAATCCAGGACTTGGAAGTGCGCTTCATTCTCCCCATCT  1740

Qy  581 LeuAspProThrMetProAlaLeuAsnTrpAsnProGlyAsnProValGlyAlaArgAlaAsp  600
Db 1741 CTTGATCCAACTATGCCAGCTCTTAACTGGAATCCCGGAAACCCCGTGAC  1800

Qy  601 GluTrpAspThrMetPheLysIleProAsnAsnTrpAsnProGlyAsnProValGlyValGlyPhe  620
Db 1801 GAGTGGGATACCATGTTCAAAATCCCGAATAACTGGAATCCTGGAGAGTGGGTTC  1860

Qy  621 LysMetAsnSerLysThrLysArgPheValGlyTyrAsnGlyValAlaAspMetProAla  640
Db 1861 AAAATGAACTCAAAAAACCAAAACGATTCGTTGGAGGAAATGGAGCATTTGATATGCCTGCA  1920

Qy  641 LeuGlyLeu  643
Db 1921 CTGGGATTG  1929
```

```
RESULT 2
US-10-028-248A-35
; Sequence 35, Application US/10028248A
; Publication No. US20030235882A1
; GENERAL INFORMATION:
; APPLICANT: Shimkets, Richard
; APPLICANT: Patturajan, Meera
; APPLICANT: Vernet, Corine
; APPLICANT: Casman, Stacie
; APPLICANT: Malyankar, Uriel
; APPLICANT: Shenoy, Suresh
; APPLICANT: Spytek, Kimberly
; APPLICANT: Gangolli, Esha
; APPLICANT: Miller, Charles
; APPLICANT: Boldog, Ferenc
; APPLICANT: Li, Li
; APPLICANT: Taupier Jr, Raymond J
; APPLICANT: Kekuda, Ramesh
; APPLICANT: Smithson, Glennda
; APPLICANT: Zerhusen, Bryan
; APPLICANT: Liu, Xiaohong
; APPLICANT: Colman, Steven
; APPLICANT: Tchernev, Velizar
; APPLICANT: Si, Jingsheng
; APPLICANT: Edinger, Shlomit
; APPLICANT: Stone, David
; APPLICANT: Sciore, Paul
; APPLICANT: Millet, Isabelle
; APPLICANT: Rothenberg, Mark
; TITLE OF INVENTION: No. US20030235882A1el Nucleic Acids and Polypeptides and Methods
; TITLE OF INVENTION: Thereof
; FILE REFERENCE: 21402-222
; CURRENT APPLICATION NUMBER: US/10/028,248A
; CURRENT FILING DATE: 2001-12-19
; PRIOR APPLICATION NUMBER: 60/256619
; PRIOR FILING DATE: 2000-12-19
; PRIOR APPLICATION NUMBER: 60/262959
; PRIOR FILING DATE: 2001-01-19
; PRIOR APPLICATION NUMBER: 60/272408
; PRIOR FILING DATE: 2001-02-28
; PRIOR APPLICATION NUMBER: 60/285189
; PRIOR FILING DATE: 2001-04-20
; PRIOR APPLICATION NUMBER: 60/308039
; PRIOR FILING DATE: 2001-07-26
; PRIOR APPLICATION NUMBER: 60/311266
; PRIOR FILING DATE: 2001-08-09
; NUMBER OF SEQ ID NOS: 211
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 35
; LENGTH: 7396
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-028-248A-35

Alignment Scores:
Pred. No.:                              Length:            7396
Score:               2.59e-06           Matches:           170
Percent Similarity:  176.50             Conservative:      95
Best Local Similarity: 31.25%           Mismatches:        256
Query Match:         20.05%             Indels:            329
Db:                  5.36%              Gaps:              34
                     17
```

US-10-736-868-2 (1-643) x US-10-028-248A-35 (1-7396)

```
Qy     12 GlyPheCysIleAlaProLeuSer------AlaGlnSerProSerThrSerAsp------ 27
          |||    |||::      ::     |||        |||||||
Db   3999 GGCTTCTCAGCCAGTCCACACAGCAAGTCCAGCAAGCAAGGACTTCTCCGGCCTGG 4058

Qy     28 AlaProGly------AlaLeuLeuSerSerLeuValGlyLysSerHisGlnLysLeuPro 45
          ::||| ||        |||         |||     |||||||||||
Db   4059 AGTCCCAGCTGCAGGACACTCAGGAGCTG-CTGCAGGAGGAGAACCCGGCAGAAGCTGAGC 4117

Qy     46 LeuAlaProSerMetGluAlaLeuGlu------ 54
          |||:::   |||||
```

```
Db   4118 CTGAGCACCAAGCTTCAAGCAGTCGAGGTGGAGCAGAGAAGAATTCCTTCCGGGAGCAGCTGGAG 4177
Qy     55 ------LeuMetGlyValGlnPheVal 61
                      :::   |||
Db   4178 GAGGAGGAGGCCAAGCACCTGAGAGCAGATCGCCACCCTCATGCCCAGGTGGCC 4237
Qy     62 AspAlaLeuIleLysGlyGlnMetGluMetAlaLysMetAlaPheLysThrGlnLeu 81
          |||        ||| |||:: |||||||
Db   4238 GAC------ATGAAAAAGATGGAGACAGTGGGGTGCCTGGAAACTGCTGAG 4288
Qy     82 GluValLeuGluLysValHisProAsp 90
          |||||:::|||
Db   4289 GAGGTGAAGAAGCTTCAGAAGGACCTTGGAGGCCTGAGCCAGGCACGAGGAGAAG 4348
Qy     91 ------GlnPheAspLysTyrLysLysLeuLys------ValAspAsp 102
                :::||||||    ::|||   |||       :::|||||
Db   4349 GTGCGGCCGCTCACGACAAGCTGGAGAAGACCAAGACGCGGCTGCAGGAGCTGACGAC 4408
Qy    103 LeuAlaLaAspAlaVaLMetGInGlnAlaLysLeuGlnMetAlaLysLeuLeuGlnProLysSerGly 122
          |||||:::      |||          |||||::
Db   4409 CTGCTGGTGGACCACCAGCGCCAGCGCGTGCAACCTGGAGAAGAAGCAGAAG 4468
Qy    123 AsnAlaPheIleAspMetLeuAsnGlyAsnGlyIleProIleGly------ 137
          |||   |||    |||
Db   4469 AAG------TTTGACCAGTCTCTGGCGAGGAGAAGAACCATCTCTGCCAAGTATGCAGAG 4522
Qy    138 ------SerSerIle 140
                |||
Db   4523 GAGCGGCCACCGGCTGAGCGCCCGAGGAGGAGGAGGAGACCAAGGCTCTGTCGCTGGCC 4582
Qy    141 ArgGlyLeuGluAspAlaIleArgThrGlnAsnThrAspMetGluAsnThrAspProSerGlu 160
          |||          |||          :::|||
Db   4583 CGGGCCCCTGAGGGAGCCCATGAACAGCAGAGAGGCCGGAGCAGGCCGGCTCAAC------ 4633
Qy    161 GlnIleAlaLysAlaValMetAspLysPheGlnThrGlnIleLeuProGlyLeuValAla 180
          |||:: |||||        :::
Db   4634 ------AAGCAGTTCCGACGGAG------------------ATGGAG 4657
Qy    181 AsnMetIleAlaGlyLysAsnProPhe------LysMetProGlnMetArgLysAlaGln 199
          |||||::::: |||          :::::     |||
Db   4658 GACCTTATGAGCTCCAAGGATGATGTGGGCAAGAGTGTCCAGAGCTGGAGAAGTC-CAA 4716
Qy    200 AlaAlaProSerValPheGlnGlnAlaLeuAlaLeuAlaMetLeuGlyLysAsn 219
          ||| ||          |||          :::
Db   4717 GCGGGGCCCT------------------------------AGAGCAGCAGCTGGAGCAGAGAT 4746
Qy    220 ------AlaProValAlaGlyValGlyArgGlyGluGluGlnLnArgMetMetAsn-Ar 236
                |||||||||:::||
Db   4747 GAAGACGCAGCTGGAAGAGCTGAGGAGGACGAGCTGCAGGCCACCGAAGATGCCAAGCTGCG 4806
Qy    236 gValAspGlnArgMetGln------GlnArgGluLeuGlnLeuGluuAs 250
          |::::: :::||||||||||::
Db   4807 GTTGGAGGTCAACCTGCAGGCCATGAACCTGAGACGCCGCAGGGCCCGGGA 4866
Qy    250 pGluAspAspAspLeuGluAspValProArgProArgArgSerAspGlyl 270
          |||||          ::: ||||      :::
Db   4867 CGACGAGCGGCGAGGAAGAAGAAGAGCAGCTGGTCAGACAGTGCGGGAGATGGGAGGCAGA 4926
Qy    270 uProGlnSerGluAlaGluHisGlnLysAspLeuAlaArgArgLeuLysSerSerPr 290
          |          |||:: ||||||       |||||| |||| |||
Db   4927 GCTGGAGGACGAGGAGGAGGAGCAGCGCTCGATGCGGCCGCTGGCGGCCGGAAGAAGCTGGAGAT 4986
Qy    290 oArgLeuLysGluLeu------ 295
          |||:::::||
Db   4987 GGACCTGAAGGAGCTGGAGGGCCACCATGCGCCCGGCCCACAAGAACCGGGACGGAAGCCAT 5046
Qy    296 ------LeuGlnAsnAlaGluValGlnSerLeuLeuSerTyrGlnArgMetArgAspSe 313
                |||::  |||:::     :::
Db   5047 CAAACAGCTGCGGGAAGCTGCGGCCAGGCGCCAGATGAAGGAACTCATGCGCGAGCTGGATGACAC 5106
Qy    313 rProLeuSerLysArgArgProLeuAlaMetAsnAspGluSerAlaPheArgAla1 333
          ||| :::: |||
```

```
Db  5107 CCGCGGCCTTCTCGTGAGGAGATCCTGGCCCCAAAGAGAGAGAGAGAGAGAAGAGCTGAAGAG 5166
Qy   333 aMetGluAla-----------------------ArgAlaAla-- 339
              ||||||||||
Db  5167 CATGGGAGGCCCGAGATGATCCAGTTGCAGGAGGAACTGGCAGCCGCCGAGCCGTGCCAAGCG 5226
Qy   339 ------------------------------------------- 339

Db  5227 CCAGGCCCCAGCAGGAGGAGGCCGGATGAGTGGCTGACAGAGATCGCCACACAGCAGCCGCAAAGG 5286
Qy   340 ------LeuAspGlnLysSerGlnLeuValLeuGlyLeuHisGlyPheGlyGluse 356
              ||||||||  :::|||
Db  5287 AGCCCTGGCGTTAGAGGAGAGGCGGGCCGTCTGGAGGACCCGGGCCCAGCTGGAGGAGGA 5346
Qy   356 rAspAspAspGluAspGluAspGluAsnLeuIleAspLysAspProSerGluAsnSerPheAr 376
              :::::|||
Db  5347 GCTGGAGGAGGAGGCAGGGCAACACGGACCTGATCAACGACCGGCTGAAGAGGAGGCCAACCT 5406

Qy   376 gArg-----------AlaProLeuArgLeuSerSerGlyPheValGluValLysLeuLy 391
              |||
Db  5407 GCAGATCGACCAGATCAACGCCCGACCTGAACCTGGAGCGGCGGCACGCCCAGAAG----- 5461

Qy   391 sSerAsnAspGluLeuLysSerAlaLeuAspArg-----------IleLysTyrArgVa 407
              ||||   :::  |||
Db  5462 ----AACGAGAATGCTCGGCACGAGCAGCTGGACAACGACAAGGAGCTTAAGGTCAAGCT 5517

Qy   407 lAspAspValGlu-----------LysTyrLeuAla-ProLysPro------- 418
              ||||  :::::|||
Db  5518 GCAGGAGGATGGGAGGGCACTGTCAAGTCCAAGTACAAGGCCTCCATCACCGCCCTCGAGGC 5577

Qy   419 --------------MetGluPheAsnProLysPro--------------GlnProG 428
              :::   |||||||
Db  5578 CAAGATTGCACAGCTGGAGGAGCAGCTGGACAACGAGCAAGGAGCGCCAGGCAGCCTG 5637

Qy   428 lyTyr------- 429
Db  5638 CAAACAGGTGCGTCGACCGAGAAGAGCTGGAAGGATGTGCTGCAGGTGGATGACGA 5697

Qy   430 ----PheAlaProArgLysIleProThrArgProArgLysMetLeuProLeuLeuIleG 448
              :::||  :::|||
Db  5698 GCGGAGGAACCGCGAGCAGCTGCGACGACCAGGCGATCTCACCCCGCTGAAGCA 5757

Qy   448 lySerAspProLysValGlnGluValIleLeuArgArgHisProSer------- 462
              :::|||  |||||
Db  5758 GCTCAAGCGGCAGCTGGAGGAGGCCG---AAGAGGAGGCCCAGCGCGGCCCACGCCTCCCG 5814

Qy   462 ------------------------------------------- 462
Db  5815 CCGGAAACTCGCAGCGCGCAGCTGGAGGACGGCCCACTGCCATGCAGGACCGGA 5874

Qy   463 ---------------------------ThrG 464
Db  5875 AGTCAGCTCCCTAAAGAACACAGCTCAGCGCGCGGGACCTGCCGTTTGTCGTGCCCGCCG 5934

Qy   464 luTrpLys-------- 472
              |||||
Db  5935 AATGGGCCCGGAAAGGCGGGGATGGCGTCCGGAAGAGTAGATGGCAAAGCCGATGG 5994

Qy   472 rgValLeuThrAsnLeuLysGlyArgGlnMetLeuThrAspGluGlnLysGlyIleArgThrArgV 492
              |||||  :::|||
Db  5995 GGCTGAGGGCCAAACCTGCCGAATAAGCCTCTCTCTCCTCCAGCCTGAGCGATGATGGACAGA 6054

Qy   492 euGluAsnThrLeuLysGlyArgLeuProArgLeuPheGlyAla----------- 512
              |||  :::
Db  6055 CAGAACCACCAGCGCCTCCCCAGG----ACCCCGCAG------ 6089

Qy   512 aLysThrIleArgAlaLysValPheGluGlnAspIleGluGlnArgProIleProP 527
              |||  :::::|||
Db  6090 CACGGCCTCCCCACCTTCTTGGGACTGCTGTGAACATGCCTCCTCCTG 6138

Qy   527 laLysAlaGlumetIleAspAlaLysValPheGlnAspIleGluGlumetLeuProProP 547
              |||
Db  6139 CCCTC---------------------CGCCCCGTCCCCC 6156
```

```
Qy   547 roLeuPhePheGluProLysLysProGlyArgHisThrArgLeuArgArgTrpThrGlyAlaAsnGluL 567
              |||
Db  6157 CATCCCGTTTCCT----------------------------------- 6170

Qy   567 ysGlyIleProGlyLeuGlySerArgPheIleLeuProSerLeuAspPro-----T 584
              |||||  :::|||
Db  6171 ----------CCAGGTGTTGTTGAGGGCATTGGGCTTCCTCCTGCATCCCCTTCCAGCTC 6222

Qy   584 hrMetProAlaLeuAsnThrAlaPheSerThrGlnGlyValArgAlaArgAspGluTrpAspT 604
              ||||||  :::|||
Db  6223 CCTCCCCTGCTCAGAATCTGATACCAAAGAGACAGGGCCCGGGCCCAGCGAGAGCCGACC 6282

Qy   604 hrMetPheLysIleProAsn 610
              |||::|
Db  6283 AGCAGGCTCCTCAGCCCTCT 6302
```

RESULT 3
US-10-107-782-35
; Sequence 35, Application US/10107782
; Publication No. US20040018970A1
; GENERAL INFORMATION:
; APPLICANT: Boldog, Ferenc,
; APPLICANT: Casman, Stacie,
, APPLICANT: Colman, Steve,
, APPLICANT: Edinger, Shlomit,
, APPLICANT: Gangolli, Esha,
, APPLICANT: Kekuda, Ramesh,
, APPLICANT: Li, Li,
, APPLICANT: Liu, Xiaohong,
, APPLICANT: Malyankar, Uriel,
, APPLICANT: Miller, Charles,
, APPLICANT: Millet, Isabelle,
, APPLICANT: Patturajan, Meera,
, APPLICANT: Rothenberg, Mark,
, APPLICANT: Sciore, Paul,
, APPLICANT: Shenoy, Suresh,
, APPLICANT: Shimkets, Richard,
, APPLICANT: Si, Jingsheng,
, APPLICANT: Smithson, Glennda,
, APPLICANT: Spytek, Kimberly,
, APPLICANT: Stone, David,
, APPLICANT: Taupier, Raymond, jr.,
, APPLICANT: Tcherney, Velizar,
, APPLICANT: Vernet, Corine,
, APPLICANT: Zerhusen, Brian
; TITLE OF INVENTION: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES AND METHODS OF USE THEREOF
; FILE REFERENCE: 21402-222CIP
; CURRENT APPLICATION NUMBER: US/10/107,782
; CURRENT FILING DATE: 2002-03-27
; PRIOR APPLICATION NUMBER: 10/028,248
; PRIOR FILING DATE: 2001-12-19
; PRIOR APPLICATION NUMBER: 60/256,619
; PRIOR FILING DATE: 2000-12-19
; PRIOR APPLICATION NUMBER: 60/262,959
; PRIOR FILING DATE: 2001-01-19
; PRIOR APPLICATION NUMBER: 60/272,408
; PRIOR FILING DATE: 2001-02-28
; PRIOR APPLICATION NUMBER: 60/285,189
; PRIOR FILING DATE: 2001-04-20
; PRIOR APPLICATION NUMBER: 60/308,039
; PRIOR FILING DATE: 2001-07-26
; PRIOR APPLICATION NUMBER: 60/311,266
; PRIOR FILING DATE: 2001-08-09
; PRIOR APPLICATION NUMBER: 60/279,344
; PRIOR FILING DATE: 2001-03-28
; NUMBER OF SEQ ID NOS: 215
; SOFTWARE: CuraSeqList version 0.1
; SEQ ID NO 35
; LENGTH: 7396
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:

```
;  NAME/KEY: CDS
;  LOCATION: (1140)..(6016)
US-10-107-782-35

Alignment Scores:
Pred. No.:              2.59e-06      Length:            7396
Score:                  176.50        Matches:            170
Percent Similarity:     31.25%        Conservative:        95
Best Local Similarity:  20.05%        Mismatches:         256
Query Match:            5.36%         Indels:             329
DB:                     17            Gaps:                34

US-10-736-868-2 (1-643)  x  US-10-107-782-35  (1-7396)
```

```
Qy      12  GlyPheCysIleAlaProLeuSer------AlaGlnSerProSerThrSerAsp------  27
                |||||  :::||       :::        |||      ||||||    ||||||
Db    3999  GGCTTCTCAGCCAGTCCGACAGCAAGTCCACCAAGGACTTCTCCGGCTGG  4058

Qy      28  AlaProGly------AlaLeuLeuSerLeuValGlyLysSerHisGlnLysLeuPro  45
                ||||            :::||||||  :::          ::::::  |||||||||
Db    4059  AGTCCGAGCTGCAGGACACTCAGGAGCTG-CTGCAGGAGGAGAACCGGAGAAGCTGAGC  4117

Qy      46  LeuAlaProSerMetGluAlaLeuGlu------------------------------  54
                |||   :::::  |||
Db    4118  CTGGAGCACCAAGCTCAAGCAGTGGAGGACGAGAAGAATTCCTTCCGGGAGCAGCTGGAG  4177

Qy      55  ------------------------------LeuMetGlyValGlnPheVal  61
                                              :::               |||
Db    4178  GAGGAGGAGCCAAGCACAACCTGAGAGCAAGTCGGAGGACGAGATCGCCCAGGTGCCC  4237

Qy      62  AspAlaLeuIleLysLysGlyMetGluMetAlaLysGlyAlaPheLysThrGlnLeu  81
                |||               ::: |||||||               :::||
Db    4238  GAC------------ATGAAAAAGAAGATGGAGCAGTGTGGGGTTGCCTGGAAACTGCTGAG  4288

Qy      82  GluValLeuGluLysValHisProAsp------------------------------  90
                |||||||
Db    4289  GAGGTGAAGAGAGAAGCTCCAGAACGACCTGGAGGGCCTGAGCCAGCGGCACGAGGAGAAG  4348

Qy      91  -----GlnPheAspLysTyrLysLysLeuLys------------------ValAspAsp  102
                     :::||||||        |||
Db    4349  GTGGCCGCCTACGACAAGTGGAGGAAGAAGAACCAAGAGCGCGGCGAGAGTGGAGGACGAG  4408

Qy     103  LeuAlaAlaAspAlaValMetGlnGlnAlaGluMetAlaLysAlaLeuGlnProLysSerGly  122
                ||| :::::  |||||::::
Db    4409  CTGCTGGTGGACGAGAAGCTCCGACCACCAAGCGCCGTGCAACCTGGAGGAGAAGCAGAAG  4468

Qy     123  AsnAlaPheIleAspMetLeuAsnGlyTyrAsnGlyIleProIleGly-----------  137
                ::::|||||||      ||:
Db    4469  AAG------TTTGACCAGCTCCTGGCAGGAGGAAGACCCATCTCTGCCAAGTATGCAGAG  4522

Qy     138  ----------------------------------------SerSerIle  140
                                                        :::
Db    4523  GAGGCGCGGCGGCCGGAGGCCCCGAGAGGAGGAGGAGAAGGCTCGTGTCGCTGGCC  4582

Qy     141  ArgGlyLeuGluAspAlaIleArgThrGlnArgAspMetGluAsnThrAspProSerGlu  160
                |||||||||:::::|||      :::::::||||
Db    4583  CGGGGCCCTGGACCTCGAGGACAGGAGCCGGAGCGGCGGCTCAAC------------  4633

Qy     161  GlnIleIleAlaLysAlaValMetAspLysPheGlnThrGlnIleLeuProGlyLeuValAla  180
                                                        :::
Db    4634  ------------------------------------------AAGCAGTTCCGCACGGAG  4657

Qy     181  AsnMetIleAlaGlyLysTyrAsnProPhe---LysMetProGlnMetArgLysAlaGln  199
                :::::::::::||:::::
Db    4658  GACCTTATGAGCTCCAAGGATGTGGGCAGAGTGTCCACGAGCTGGAGAAGTC-CAA  4716

Qy     200  AlaAlaProSerSerValPheGlnGlnIleAlaLeuAlaMetLeuGlyLysLysAsn  219
                |||  |||                      ::    :::
Db    4717  GCGGGGCCCT------------AAGAGGAGGCCCCAACCGCCTCCCG  4746

Qy     220  ---AlaProValAlaLaGlyArgGlyArgGlyGlnGlnMetMetAsn-Ar  236
                   ||| |||
Db    4747  GAAGACCGCAGCTCGAAGAGACGAGCTCGAGGAGGAGGCCACCGAAGAGATGCCAAGCTGCG  4806
```

```
Qy     236  gValAspGlnArgMetGln------  250
                |||||
Db    4807  GTTGGAGGTCAACCTGCAGCGCCATCAAGGCCCAGTTCGACGTGGGGCGAGGGGCGGGA  4866

Qy     250  pGluAspAspAspLeuGlyLysGluAspGlyVal------ProArgArgSerSerGlyGly  270
                |:::::::  :::::::  :::      :::  |||
Db    4867  CGAGCAGCTGGAGGAGAAGAAGAAGCAGCTGGTCGACACAGTGGCGGGAGATGGAGCAGA  4926

Qy     270  uProGlnSerGluAlaGluHisGlnHisGlnAlaArgArgAspLeuAlaArgLeuLysSerSerPr  290
                |||    :::  :::        :::::||
Db    4927  GCTGGAGGACGAGGAGGAGGCGCTCGATGGGCGGCGGGGCGGAAGAGCTGAGAT  4986

Qy     290  oArgLeuLysGluLeu------------------------------  295
                ::||||||:::||
Db    4987  GGACCTGAAGGACCTGAGGCGGCCACATCGGCCAACAGAACCGGACGAGCCAT  5046

Qy     296  ------LeuGlnAsnAlaGluValGlnSerLeuLeuSerTyrGlnArgMetArgAspSe  313
                      |||             ::: |||
Db    5047  CAAACAGCTGCAGGGAAGCTCAGGACGCTGGAGCTGGATGACAC  5106

Qy     313  rProLeuSerLysArgArgProLeuAlaMetAsnAspGluSerAlaPheArgAl  333
                |||||||:::::  ||| |||||||
Db    5107  CCGGCGCCTCTTCGTGAGGAGGATCCTGGCCCAGGCCCAAAGAACGAGAAGAAGTGAAGAG  5166

Qy     333  aMetGluAla---------------------ArgAlaLys---------  339
                :|||||||                             ||
Db    5167  CATGGAGGCCGAGATGATCCAGTTGCAGGAGGAGGAACTGGCAGCCGGCGGTGCCGAGCG  5226

Qy     339  ------------------------------------------  339
Db    5227  CCAGGCCCAGGAGGCGGGATGAGCTGCTGACGAGATCGCCCAACAGCAGCGGCAAAGG  5286

Qy     340  ------LeuAspGlnLysSerGlnLeuValLeuGlyLeuHisGlyPheGlyGluse  356
                      :::               |||
Db    5287  AGCCCTGGGCTTTAGAGGAGGAAGCGGTCTTGGAGGCCCCATCGCCCAGCTGGAGGAGGA  5346

Qy     356  rArgAspAspGluAspGlyGluAspGlyAsnLeuIleAspProSerGluAsnSerPheAr  376
                |||||:::                   ||||||
Db    5347  GCTGGAGGAGGAGGCAGGGCAACACGAGCTGATCAACGACCGGGTGAAGAGGGCCAACCT  5406

Qy     376  gArg---------------------AlaProLeuArgLeuSerSerGlyPheValGluLysLeuLy  391
                :::                       |||
Db    5407  GCAGATCGACCAGATCAACGCCGACCTGAACCTGGAGCGCGGCGGCACGCCCAGAAG-----  5461

Qy     391  sSerAsnAspGluLeuLysSerAlaLeuAspArg-------------IleLysTyrArgVa  407
                ||||                          :::::
Db    5462  ----AACGAGAATGCTCGGCAGCAGTCGGAACGCCCAGAACAAGGAGCTTAAGGTCAAGCT  5517

Qy     407  lAspAspValGlu------------------------------LysTyrLeuAla-ProLysPro  418
                :                                         :::||||||
Db    5518  GCAGGAGGATGAGGGGCACTGTCAAGTCAAGTACCAAGGGCCCTCCATCACCGCCCTCGAGGC  5577

Qy     419  ------------------------------MetGluPheAsnProLysPro  428
                                              |||
Db    5578  CAAGATTGCACAGCTGGAGGAGACTGGACAACGACCAAGAGACGCCAGCCGAGCCTG  5637

Qy     428  lyTyr------------------------------------------  429
                :::
Db    5638  CAAACAGTGGCGTCGGCACCGGAGGAGAGCTGAGGAGATGTGCTGCTGCAGGTGGATGACGA  5697

Qy     430  ------PheAlaProArgLysIleProThrArgProArgLysMetLeuProLeuLeuIleGlu  448
                      |||
Db    5698  GCCGGAGGAACGCCCAGCAGTACAAGGACCGGACACAAGGCATCTTCACCCGCCTGAAGCA  5757

Qy     448  lySerAspProLysValGlnGluValIleGluArgHisProSer------  462
                |||
Db    5758  GCTCAAGCGGCCAGCTGAGGAGGCCCG-----AAGAGGAGGCCCCAACCGCCTCCCG  5814

Qy     462  ------------------------------------------  462
Db    5815  CCGGAAACTGCACGCGGCCACTGAGGACGCGCCACTGAGAGCCATGACGAGATGCCAAGCTGCG  5874
```

```
Best Local Similarity:   20.32%     Mismatches:  263
Query Match:              5.27%      Indels:      346
DB:                       21         Gaps:        35

US-10-736-868-2 (1-643)  x  US-10-696-909A-48  (1-7274)

Qy      12  GlyPheCysIleAlaProLeuSer-----AlaGlnSerThrSerAsp-------  27
                ||| :::  |||            |||    |||||||::
Db    3860  GGCTTCTCAGCAGTCGACAGCAAGAGTCACCAAGGACTTCTCCGCCTGG  3919

Qy      28  AlaProGly-------AlaLeuLeuSerLeuValGlyLysSerHisGlnLysLeuPro  45
            :::|                 |:::  |  ||| ::::::|||||||||||
Db    3920  AGTCCCAGCTGCAGGACACTCCAGGAGCTG-CTGCAGCAGGAGGAACCCGGACAGCTGAGC  3978

Qy      46  LeuAlaProSerMetGluAlaLeuGlu--------------------  54
                |||::  ||||::  ::||:||
Db    3979  CTGGACCACCAAGCTCAAGCAGGTGGAGGACGAGAAGAATTCCTTCCGGGAGCCAGCTGGAG  4038

Qy      55  ------------------------LeuMetGlyValGlnPhe  60
                                        :::
Db    4039  GAGGAGGAGGAGCCAAGCACAACCTCACGAGAGCAGATCGCGATGGCCATGCCCCAGGTG  4098

Qy      61  ValAspAlaLeuIleLysLysGlyGlnMetGluMetAlaLysPheLysThrGln  80
                 |||:::||||::     ||||:||||||||||
Db    4099  GCCGAC------ATGGAAAAAGAAGATGAGGACAGTGTGGGGGTGCCTGAAAACTGCT  4149

Qy      81  LeuGluValGluLysValHisProAsp-------  90
            |||    ||||:  :::||||
Db    4150  GAGGAGGTGAAGAGAAGCTCCAGAAGGACTTGGAGCTG  4209

Qy      91  ---------GlnPheAspLysTyrLysLysLeuLys---------ValAsp  101
                     :::|||||||               ::::|||
Db    4210  AAGGTGGCCGTACGACAAGCTGGAGGAAGAGCGCGTCGACGAGGAGCTGGAC  4269

Qy     102  AspLeuAlaAlaAspAlaValMetGlnAlaAlaGluMetAlaLysLeuGlnProLysSer  121
                ||| :::                      ||||:::
Db    4270  GACCTGCTGGTGGACCTGGACTGGACCACCAGAGCCGTGCAACCTGGAAGAAGAAGCAG  4329

Qy     122  GlyAsnAlaPheIleAspAspMetLeuAsnGlyAsnGlyIleLeuProIleGly-------  137
                |||
Db    4330  AAGAAG------TTTGACCAGCTCCTGGGCGAGGAGAAGACCATCTCTGCCAAGTATGCA  4383

Qy     138  ----------------------------------------------SerSer  139
                                                          |||
Db    4384  GAGGAGGAGCGGCGGGCTGAGGCGGAGCGCCCGAGAAGAGGAGACCAAGGCTCTGTCGCTG  4443

Qy     140  IleArgGlyLeuGluAspAlaIleArgThrGlnAspMetGluAsnThrArgAspProSer  159
                :::|||||:::||||            :::  ||||
Db    4444  GCCCGGGCCCTGGAGGAGGCCATCAGACAGCAGGACATGGAGAACACGCGGGACCCGGCGCTCAAC-----  4497

Qy     160  GluGlnIleAlaLysAlaAlaValMetAspLysPheGlnThrGlnIleLeuProGlyLeuVal  179
                |||::  ||:::  :::|||||:
Db    4498  -----AAGCAGTTCCGACGGGAG--------------ATG  4518

Qy     180  AlaAsnMetIleAlaGlyLysGlyGluGlnAlaArgLysAla  198
                :::::|||||:::       |||
Db    4519  GAGGAGCCTTATGAGCTCCAAGATGATGGGCAAGAAGTCTCCAAGAGCTGAGAAGTC-  4577

Qy     199  GlnAlaAlaProSerSerValPheGlnAlaLeuAlaLeuGlnArgAlaMetLeuGlyLys  218
                |||:::|||||||||       |||
Db    4578  GCGGTTGGAGGTCAACCTGCAGCCCATGAAGCCCCAGTTCGAGCCGGACCTGAGCGGGCCG  4607

Qy     219  Asn-------AlaProValAlaGlyGlyArgGlyGluGlnLeuArgMetMetAsn  235
               :::       ::: |||::
Db    4608  GATGAAGAGACGCAGCTGGAAGAGCTGGAGGACGAGCTGCAGGACGCTGCCAAGATGCCAAGCT  4667

Qy     236  -ArgValAspGlnLeuGlnLeuGlu1  249
             :::||||:::||||||
Db    4668  GCGGTTGGAGGTCAACCTGCAGGCCCATGAAGGCCCCAGTTCGAGCCGGACCTGAGCGGGCCG  4727

Qy     249  uAspGluAspAspAspLeuGluAspGlyGluGluGluuAlaGlyValProArgArgSerAspG1  269
                :::|||  :::  :::::      ::
Db    4728  GGACCGAGAGCGGAGGAGGAGGAAGAGAAGGAGCAGTGGTCAGACAGGTGCGGGGAGATGGGAGC  4787
```

```
Qy     463  -----------------------ThrG  464
Db    5875  AGTCAGCTCCCTAAAGAACAAGCTCAGGCGGGGACCTGCCGTTGTCGTCGCCCCGCCG  5934

Qy     464  luTrpLys-----------------------IleAlaLysGluSerA  472
                |||||                           :::|||||||
Db    5935  AATGGCCCGGAAAGGCCCGGGATGGCTCCGACGAGAGGAGCTAGATGGGCAAAGCGGATGG  5994

Qy     472  rgValLeuThrAsnLeuLysAsnAsnProSerLeuLeuAlaLeuPheMetAspAspLysL  492
                ||| :::  ||||  :::
Db    5995  GGCTGAGGCCAAACCTGCCGAATAAGCCTCTTCTCCTGCCAGCCTGAGCCTGAGCCTGAGATGGACAGA  6054

Qy     492  euGluAsnThrLeuLysGlyArgGlnMetLeuThrAspGluGlnLysGlyArgThrArgV  512
                :::  |||||||       ||||  :::
Db    6055  CAGACACCACCAGCCTCCCCCTTCCAG-------ACCCCGCAG  6089

Qy     512  alLysThrIleArgAlaLeuProArgLeuPheGlyAla------------ProThrA  527
                ||| |||   |||  |||
Db    6090  --------CACGCCTCTCCCCACCTTCTGCGGACTGCTGTGAACATGCCTCCTCCTG  6138

Qy     527  laLysAlaGluMetIleAspAlaLysLysValPheGlnAspLeuGluGlnArgProP  547
               |
Db    6139  CCCTC------------CGCCCGTCCCCC  6156

Qy     547  roLeuPhePheGluProLysGlyLysArgHisThrArgTrpThrGlyAlaAsnGluL  567
                ||||||
Db    6157  CATCCCGTTTCCCT-------------  6170

Qy     567  ysGluIleProGlyLeuGlySerArgPheIleLeuProSerArgLeuAsp-------T  584
            |||||||::                   :::
Db    6171  ------CCAGGTGTTGTTGAGGGCATTTGGCTTCCTCCTGCTCCCCCTTTCCAGCTC  6222

Qy     584  hrMetProAlaLeuAsnThrAlaPheSerThrGlnGlyArgAlaArgAspGluTrpAspT  604
                |||||||||
Db    6223  CCTCCCCTGCTCAGAATCTGATACCCAAAGAACAGAGACACGGGCCCCGGGCCCAGGCCAGCAGAGCGACC  6282

Qy     604  hrMetPheLysIleProAsn  610
                |||:::
Db    6283  AGCAGGCTCCTCAGCCCTCT  6302
```

RESULT 4
US-10-696-909A-48

; Sequence 48, Application US/10696909A
; Publication No. US20050118604A1
GENERAL INFORMATION:
; APPLICANT: Lorens, James B.
; APPLICANT: Atchison, Robert E.
; APPLICANT: Friera, Anabella
; APPLICANT: Holland, Sacha
; APPLICANT: Rigel Pharmaceuticals, Inc.
; TITLE OF INVENTION: Modulators of Angiogenesis and Tumorigenesis
; FILE REFERENCE: 021044-005820US
; CURRENT APPLICATION NUMBER: US/10/696,909A
; CURRENT FILING DATE: 2003-10-29
; PRIOR APPLICATION NUMBER: US 60/512,251
; PRIOR FILING DATE: 2003-10-17
; PRIOR APPLICATION NUMBER: US 60/421,989
; PRIOR FILING DATE: 2002-10-29
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 48
; LENGTH: 7274
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: myosin, heavy polypeptide 9, non-muscle (MYH9)
; OTHER INFORMATION: cDNA
US-10-696-909A-48

Alignment Scores:
Pred. No.:   5.03e-06   Length: 7274
Score: 173.50   Matches: 178
Percent Similarity: 30.71%   Conservative: 91

```
Qy   269  yGluProGlnSerGluAlaGluHisGlnArgArgAspLeuAlaArgArgLeuLysSerSe  289
Db  4788  AGAGCTGGAGGACGAGGAAGCAGCAGCGCTCGATGGCAGTGCCCCCGGAAGAGAGCTGGA 4847

Qy   289  rProArgLeuLysGluLeu-----------------------------------------  295
Db  4848  GATGGACCTGAAGGACGCACATGCTCGGCCACAACAGACCGGGACGAAGC           4907

Qy   296  -------LeuGlnAsnAlaGluValGlnSerLeuLeuSerTyrGlnArgMetArgAs     312
Db  4908  CATCAAACAGCTGCGGGAGCTGCAGGACGCCAGGATGAAGGACTGCATGCGCGAGCTGGATGA 4967

Qy   312  pSerProLeuSerLysArgArgProLeuAlaMetAsnAspGluAspGluSerAlaPheAr  332
Db  4968  CACCCGGCCGCTCGTGAGGAGATCCTGCCCAGGCCCAAAGAGAACGAGAAGAAGCTGAA  5027

Qy   332  gAlaMetGluAlaArgAlaLysLeuAspGlnLysSerGlnLeuValLeuGlyLeuHisGl  352
Db  5028  GAGCATGGAGGCC----------GAGATGATCCAGTTGCAGGAGGAACTGGACAGC      5072

Qy   352  yPheGlyGluSerAspAspAspGluGlnAspGluGluAsnLeuIleLeuAspProSerGl  372
Db  5073  CGCGGAGGCGTGCCAAGCGCCAGGCCAGGCGGGATGAGCGGCTGGCTGACGACGATCGC  5132

Qy   372  uAsnSerPheArgArgAlaProLeuArgLeuSer--------------------------  383
Db  5133  CAACAGCAGCGCGCAAAGGAGCCCTGGCGTTAGAGGAGAGCGGCGTCTGGAGGCCCGCAT 5192

Qy   384  -SerGlyPheValGlyLysLeuLysSer-------------------------------  395
Db  5193  CGCCAGCTGAGGAGGAGGACTGGAGGAGGAGGCAGGGCAACACGGAGCTGATCAACGACCG 5252

Qy   395  uLeuLysSerAla----------------------------------------------  399
Db  5253  GCTGAAGAAGGCCAACCTGCAGATCGACCAGATCAACACCGACCTGAGACCTGAGACGGCAG 5312

Qy   400  ----------------LeuAspArg---------------------------IleLeu   404
Db  5313  CCACGCCCAGAAGAACAGAATGCTCGGCAGCAGCTGGAACCGCCAGAACAAGGAGCTTAA  5372

Qy   404  sTyrArgValAspAspValGlu------LysTyrLeuAla-ProLysPro-          418
Db  5373  GGTCAAGCTGCAGGAGATGGAGGGCACTGTCAAGTACAAGGCCTCCATCACCGC        5432

Qy   419  ----------------MetGluPheAsnProLysPro-----------------       425
Db  5433  CCTCGAGGCCCAGATTGCACAGCTGGAGGAGCAGCTGGACAACAACGAGGAGCGCCA     5492

Qy   426  --GlnProGlyTyr---------------------------------------------  429
Db  5493  GGCAGCCTGCAAACAGGTGCGTCGGACCAGAGAAGAGAGTGAAGGATGTGCTGCTGCAGGT 5552

Qy   430  -----------PheAlaProArgLysIleProThrArgLysMetLeuProL          445
Db  5553  GGATGACGAGCGGAGGAACGCCGAGCAGTACAAGGCCAGGCCGACAAGGCATCTACCCG   5612

Qy   445  euLeuIleGlySerAspProLysValGlnGlnValGlnIleLeuArgHisProSer----- 462
Db  5613  CCTGAAGCAGCAGTCCAAGCGCAGCAGCTGGAGGAGGCCG----AAGAGGAGCCCCAGCGCGGGGCCAA 5669

Qy   462  ----------------------------------------------------------- 462
Db  5670  CGCCTCCCGCGCCGGAAACTCAGCGCGAGCTGGAGCACCACCCCACTGAGACGCCGATGCCAT 5729

Qy   462  ----------------------------------------------------------- 462
Db  5730  GAACCGGAAGTCAGTCTCCCTAAAGAACAAGTCAGGCGCGGGGACCTGCGCGTTTGTCGT 5789

Qy   463  ----ThrGluTrpLys-----------------------------IleAlaL        469
Db  5790  GCCCCGCCGGAATGGCCCGGAAAGGCGCCGGGGGATGGCTCCGACGAAGAGGTAGATGGCAA 5849
```

```
Qy   469  yGluSerArgValLeuThrAsnLeuLysAsnProSerLeuAlaAlaLeuPheMetA     489
Db  5850  AGGGATGGGGCTGAGGCGAACCTGCCAAACCTGCGGAATAAGCCTCTTCCTCGAGCCTGAGATGG 5909

Qy   489  spAspLysLeuGluAsnThrLeuLysGlyArgGlnMetLeuThrAspGlnLysGlyLysA  509
Db  5910  ATGGACAGACAGACACCACAGCCTCCCTTCCCAG------ACCCCGCAG------       5953

Qy   509  rgThrArgValLysThrLeuArgAlaLeuProArgLeuPheGlyAla-             524
Db  5954  ------CACGGCCTCTCCCCACCTTCTTGGGACTGCTGTGAACATGC              5993

Qy   525  --ProThrAlaLysLysAspIleAspAlaLysAlaValPheGlnAspIleGluGluArgP  544
Db  5994  CTCCTCCTGCCCTC--------------------------------CGCC          6011

Qy   544  roIleProProLeuPhePheGluProLysGlyArgHisThrArgLeuArgTrpThrGlyA  564
Db  6012  CCGTCCCCCATCCCGTTTCCCT--------------------------            6034

Qy   564  laAsnGlyLysGlyIleLeuProGlyLeuGlySerArgPheIleLeuProSerLeuAspPro- 583
Db  6035  --------------------CCAGGTGTTGTTGAGGGCATTTGGCTTCCTCTGCTGCATCCCC 6077

Qy   584  --------ThrMetProAlaLeuAsnThrAlaPheSerThrGlnGly-ArgAlaArgAsp  600
Db  6078  TTCCAGCTCCCTCCCCTGCTCAGAATCTGATACCAAAGAGAACAGGGCCCGGGCCCAGGCA 6137

Qy   601  GluTrpAspThrMetPheLysIleProAsn------------------------------ 610
Db  6138  GAGAGCGACCAGGAGCCCTCTCCAGCCCTTCTTCTTGCCAAAAGCACAAGAGATGTTGAGGCGA 6197

Qy   611  --AsnTrpAsnProGlyAspGluValGlyPheLysMetAsn                    623
Db  6198  GCAGGGCAGGGCCCCGGGGAGGGGCCAGAGTTTTCTATGAAT                  6239
```

```
Alignment Scores:
Pred. No.:
Score:                   3.44e-06       Length:        2791
Percent Similarity:      169.50         Matches:        140
Best Local Similarity:   37.73%         Conservative:   103
Query Match:             21.74%         Mismatches:     226
DB:                      5.15%          Indels:         177
                         20             Gaps:            29

US-10-736-868-2 (1-643) x US-10-425-115-63082 (1-2791)

Qy    20  AlaGlnSerProSerThrSerAspAlaProGlyAlaLeuLeuSerSer------------  35
Db    76  GCAGCGCCGCGGCCGCGCCGGGGGATGGCTCCGACGAAGAGGTAGATGGCATC        135
```

```
Qy    36  -----------------LeuValGlyLysSerHisGlnLysLeuProLeuAla   47
                           :::||||||||||||
Db   136  CCGATCCCCTACGATCGCTGAGGAGCATTATGGGTAAATCAAGCAAGAGTCCAGTGTTAAG  195

Qy    48  ProSerMetGluAlaLeuGlnLeuMetGlyValGlnPheValGlnPheAlaLeuIleLysLys   67
             |||||                 :::                        |||||||
Db   196  GCTGCACCTGCTGCTGTCTCAGTTCCAGAGGGAAATCTGCA------------AAGAAG  243

Qy    68  Gly--------------GlnMetGluMetAlaLysGlyAlaPheLysThrGlnLeu   81
            :::||                 :::  |||
Db   244  GGGAAGAGAAATGCAGAAGATGAGAATTGAGAAGGCCCTGAGTGCA---AAGAAGCAAAAG  300

Qy    82  GluValLeuGluLysValHisProAspLysTyrLysGlyLeuLysValLeuLysValAsp  101
          |||||||||||||||:::                  ||||||||
Db   301  ACTGTACCTGCTGCTGTGCGGTTAAGGAGCAGCCCCCTCCAAAG--AAAGTTGAG  357

Qy   102  AspLeuAlaAlaAspAlaValMetGlnGlnAlaGluMet---AlaLysLeuGlnProLys  120
             :::::::::::::                            |||::::||||||
Db   358  AGCAGCAGCCTCTGAGGAGGATTCTTCGAGTCCGAGGTGAGGTTAAGGTCCAGCCGAAG  417

Qy   121  SerGlyAsnAlaPheIleAspMetLeuAsnGlyAsnGlyIleProIleGlySerSerIle  140
                                                    ||||||||||||
Db   418  AAG--------------------------------------GTTATTCAACCGAAG  435

Qy   141  ArgGlyLeuGluAspAlaIleArgThrGlnArgAspMetGluAsnThrAspProSer-  159
             :::  :::                         |||  :::  |||
Db   436  AAAGGTACACACCTCTAAACAGAGTCCAGTGATGACAGCACCTGGAGAGCTCCTCA  495

Qy   160  ---GluGlnIleAlaLysAlaValMetAspLysPheGlnThrGlnIleLeuProGlyLeu  178
             :::||||            |||::::::||||||
Db   496  GACGATGAACCTTCCAAAAAACCAGTTGC-CGGATTCAAAAAAGCCTCCTGCCAGTAGTA  554

Qy   179  ValAlaAsnMetIleLeuAlaGlyLysLysAsnProPheLysMetProGlnGlnMetArgLysAla  198
                    ||||||                    |||  :::
Db   555  GCAGCAGTGATGATAGCA-----------GTGACAGTGATGATAGCAGCTCTGATGAGGACCAG  608

Qy   199  GlnAlaAlaProSerValPheGlnGln-----------AlaLeuAlaGlnArgAla  214
             :::||                  ||||||
Db   609  CAAGGAAGCCCACGGTCC---TTTCAAAGGCTCCACGGTTCAAAAGAGCCCCACCT  665

Qy   215  MetLeuGlyLysLysAsnAlaProValAlaGlyLyrGlyGluGln---------------  230
             |||                        |||||||||
Db   666  GTCCTTTCAAAAAAACCTGCCAAAAAACCTGGTTCAAACAGGTTAAACCTGAT  725

Qy   231  ----------------------------------Arg  231
                                               :::
Db   726  AGCAGCAGCTCTGACAGTAGCTCAGATGAGGATGAGAAAAACCTGCTGCTAAGGAGAG  785

Qy   232  MetMetAsnArgValAspAspArgMetGlnArgMetGlnLeuGlnMetGluGlnAspGlu  251
             :::::|||                |||  :::
Db   786  CCTGTTGTTGCCTGTGCAAAAGAGACCCAAGAATCTGATAGCTGACAGTGATTCT  845

Qy   252  AspAspAspAspProLeuValAspGluGluAspValProArgArg  264
             :::|||||::::                            :::
Db   846  GATGAGGATGAATCAGAAGAT--GATATTCCTGCTAAAGCTCAAGTAGCCAAGAGAA  902

Qy   265  -----------ArgSerSerAspGlyProGlnSerGluAlaGlnHisGlnArg  279
                         ::::::||||||||::::||
Db   903  AAGGAAGAGAGTCCAGTAGCAGTTCTGATTCTGATAGTGATTCCGAGTGAGAGAAA  962

Qy   280  ArgAspProAlaArgLeuLysSerSerProArgLeuLysGluLeuGlnAsnAla  299
             |||||||                 |||||||
Db   963  CCT---GCTGCTAACCTAAAGAGAGCCTCCG--------------------  989

Qy   300  GluValGlnSerLeuLeuSerTyrArgAsnMetArgAspSerProLeuSerLysArgArg  319
             ||||||                                         |||||||
Db   990  ---GTTCCCTCTGTGCAAAAGAGAGCCCAAGAATCTGACAGTGAT------  1040

Qy   320  ProLeuAlaMetAsnAspGluLeuAspGluSerAlaPheArgAlaMetGluArgAlaLys  339
             :::||::::||||:::::||::::||            |||||::||||
Db  1041  --------------TCTGATGAGGATGAATCA-----GATGATATTCCTGCTAAAGCTCAA  1082

Qy   340  LeuAspGlnLysSerGlnLeuValLeuGlyLysHisGlyPheGlyGluSerAspAspAsp  359
             |||
Db       LeuValGlyLysSerHisGlnLysLeuProLeuAla (see top)
```

```
Db  1083  GTAGTAGCCAAGAAAAGGAAGAAGAATCTGGTAGCAGTTCTGATTCTGATAGTGAATTCGAG  1142

Qy        ---------------LeuIleAspPro  370
                          |||
Db  1143  TCTGACGATGAGGATAGCAATGCTAAAACAGTACAACCTGCTAAAGCCTGCTGCTATTAAA  1202

Qy        ----------------------------SerGluAsnSer  374
Db  1203  AAGAAGGAAGAGATTCTAGTAGCAGTTGAGAGTCCAGAAACTGACTCCACAGTCAGAAGATTCA  1262

Qy   375  PheArgArgAlaPro-LeuArgLeuSerSerGlyPheValGluLysGluLysSerAsnAs  394
                                     ||||
Db  1263  GAACCTGAAAAACCTACTGTCCCTGCGAAAAGGCCTCAGCAACAATAAAAGAATGAA  1322

Qy   394  p----------------GluLeuLysSerAlaLeuAspArgArgIleLysTyrArgValAspAspVa  410
                              ||||||
Db  1323  CAATCCAGCGAGAAATCTGATGACAGCTCTGATGAGAGTTCTGATGAAAGTG-----  1374

Qy        ---------------------------------------------  1374
Db  1375  ----ATGAGGAACCTCCACAAAAGAAAATAAGAAATTCTGACCTTCTGGTG-----  1422

Qy   430  eAlaProArgLysIleProThrArgProArgLysMetLeuProLeuLeuIleGlySerAs  450
                                               :::||||||
Db  1423  ----CTGCAAAGGCTGCTGCCACTAAATGTTCAAAGA-------AAAA  1457

Qy   450  pProLysValGlnGluGluIleLeuArgHisProSerThrGluTrpLysProLysLysAlaLysGl  470
             :::|||::::                   |||||||
Db  1458  AGCAGCAGTGATGAAGATGAAAGTGAAAGTGTCAGAAGAAA---------GCTCTGATGATG--  1506

Qy   470  uSerArgValLeuThrAsnLeuLysAsnAsnProSerLeuAlaLeuPheMetAspAsnPro  490
             ::::::|||                    |||
Db  1507  -AGGATACAGAACAAGTTGAAACTAAGGAAACTTCTGTAGTCACTACCCAGAAGCCCG---  1563

Qy   490  pLysLeuGluAsnThrLeuGlyLysArgGlnMetLeuThrAspGluGlnLysGlyArgTh  510
                                                   |||||||
Db  1564  -AGGATATGAACCTAAAACTCCTGCCAAAAGCCAA-AGTCAGGCTACTGGTTCAAAGAC  1621

Qy   510  rArgValLysThrIleArgArgLeuProArgLeuPheGlyAlaProThrAlaLysAlaGl  530
             :::                  |||||||
Db  1622  -CACATAATGAACCTAAAACTCCTGCCAAAAGCCAA-AGTCAGGCTACTGGTTCAAAGAC  1678

Qy   530  uMetIleAspAlaLysValPheGlnAspIleGluProProLeuPhePh  550
                  :::                       ||||||
Db  1679  TATTTTTGTTGGGAATTTAGCCTACAGTATAGAGCGTGAACAAGTTAAGGAA---TTTTT  1738

Qy   550  eGluProLysGlyArgHisThrArgLeuArgTrpThrGlyAlaAsnGluLysGluIlePr  570
             |||||||                          |||
Db       TGAGGAGCTGGTGAAGTTGTTGATGTTCGTTTGGCTACTTTTGATGATGGGAGTTTCAA

Qy   570  oGlyLeuGly  573
             |||
Db       GGGTTTTGGA  1748
```

```
RESULT 6
US-10-188-248-25
; Sequence 25, Application US/10188248
; Publication No. US20040029790A1
; GENERAL INFORMATION:
```

; APPLICANT: Patturajan, Meera
; APPLICANT: Gerlach, Valerie
; APPLICANT: Anderson, David W.
; APPLICANT: Taupier Jr., Raymond J.
; APPLICANT: Zerhusen, Bryan D.
; APPLICANT: Guo, Xiaojia Sasha
; APPLICANT: Casman, Stacie J.
; APPLICANT: Hjalt, Tord
; APPLICANT: Miller, Charles E.
; APPLICANT: Kekuda, Ramesh
; APPLICANT: Shimkets, Richard A.
; APPLICANT: Malyankar, Uriel M.
; APPLICANT: Zhong, Mei

```
; APPLICANT:    Padigaru, Muralidhara
; APPLICANT:    Li, Li
; APPLICANT:    Shenoy, Suresh G.
; APPLICANT:    Gorman, Linda
; APPLICANT:    Edinger, Shlomit R.
; TITLE OF INVENTION: NOVEL HUMAN PROTEINS, POLYNUCLEOTIDES ENCODING THEM AND METHODS O
; TITLE OF INVENTION: THE SAME
; FILE REFERENCE: 21402-297D
; CURRENT APPLICATION NUMBER: US/10/188,248
; CURRENT FILING DATE:  2002-07-02
; PRIOR APPLICATION NUMBER: 60/303,046
; PRIOR FILING DATE: 2001-07-05
; PRIOR APPLICATION NUMBER: 60/303,828
; PRIOR FILING DATE: 2001-07-09
; PRIOR APPLICATION NUMBER: 60/358,932
; PRIOR FILING DATE: 2002-02-22
; PRIOR APPLICATION NUMBER: 60/304,502
; PRIOR FILING DATE: 2001-07-11
; PRIOR APPLICATION NUMBER: 60/305,011
; PRIOR FILING DATE: 2001-07-12
; PRIOR APPLICATION NUMBER: 60/305,262
; PRIOR FILING DATE: 2001-07-13
; PRIOR APPLICATION NUMBER: 60/307,536
; PRIOR FILING DATE: 2001-07-24
; PRIOR APPLICATION NUMBER: 60/306,085
; PRIOR FILING DATE: 2001-07-17
; PRIOR APPLICATION NUMBER: 60/308,228
; PRIOR FILING DATE: 2001-07-27
; PRIOR APPLICATION NUMBER: 60/323,449
; PRIOR FILING DATE: 2001-09-19
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 234
; SOFTWARE: CuraSeqlist version 0.1
; SEQ ID NO 25
; LENGTH: 6126
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (233)..(3103)
US-10-188-248-25

Alignment Scores:
Pred. No.:                             Length:           6126
Score:               169.50            Matches:          118
Percent Similarity:  36.28%            Conservative:     108
Best Local Similarity: 18.94%          Mismatches:       248
Query Match:         5.15%             Indels:           149
DB:                                    Gaps:             22
                                       17
```

US-10-736-868-2 (1-643) x US-10-188-248-25 (1-6126)

```
Qy      36 LeuValGlyLysSerHisGlnLysLeuProLeuAlaProSerMetGluAlaLeuGluLeu   55
              ||||||||                                          ||||||
Db     602 CTGACTGGCTCATCCATCATCCAACCACCACCAGGTCCCCTCCATG---TTGAGGCAGGTA   658

Qy      56 MetGlyValGlnPheValAspAlaLeuIleLysLysGlyGlnMetGluMetAlaLysGly   75
              |||                                          :::::::::::
Db     659 AGAGACAGCACAATGTTAGATCTTCAGGCCCAGCTGAAAGAACTGCAGAGAGAAATGAC   718

Qy      76 AlaPheTyrThrGlnLeuGluValGluLysValLeuGluGlyValHisProAspPheAspTyr   95
              :::    :::||||||:::::      :::
Db     719 CTCCTCCGAAAGCTAGACATCCAAGGAGCAGCAAATGGGATCTTCCATGACAGTATT   778

Qy      96 LysLysLeuLysValAspAspLeuAlaAlaAspAlaValMetGlnGlnAlaGluMetAla   115
              |||||||                                          ||||||
Db     779 AAGACTTTCTGGAGTCCTGAGCTTAAGAACGAGAGAGTCTTGAGGAAAGAGAGGCAGCG   838

Qy     116 LysLeuGlnProLysSerGlyAsnAlaPheIleAspMetLeuAsnGlyAsnGlyIlePro   135
              :::::::::::
Db     839 CGGATGTGTCCTCAAGGAGCAGATGAGGGGTTCCCATGAAGAAATCAGCACCTACAG   898

Qy     136 IleGlySerIleArgGlyLeuGluAspAlaIleArgThrGlnIleArgAspMetGluAsn   155
              :::::::::::                                          |||
```

```
Db     899 ::::|::::      :::|:::||                    :::|||||||:::      :::
           TTG---ACAATCCAGGCCCTCAAGATGAGCTGCGAACCCAGAGAGACCTCAACCAC   952

Qy     156 ThrAspProSerGluGlnIleAlaLysAlaValMetAspLysPheGlnIleLeuLeu   175
              |||                                          :::      :::
Db     953 CTCCCTCCAGCAAGAGAGTGCAACCGAGGACGG---GAGCACTTCACCATCGAGCTGACC   1009

Qy     176 ProGlyLeuValAlaAsnMetIleAlaGlyLysAsnProPheLysMetProGlnMet   195
                                                          |||
Db    1010 ------------------------------GAGGAGAACTTTAGGCGGCTCCAAGCCGAG   1039

Qy     196 ArgLysAlaGlnAlaAlaAlaProSerValPheGlnAlaLeuAlaLeuAlaArgAlaMet   215
              :::      :::||||||      :::      :::
Db    1040 CATGACAGGCAGGCTAAGGAGCTGTTCCTTTGAGGAAGACATTAGGAGGAAATGGAGCTG   1099

Qy     216 LeuGlyLysAsnAlaProValAlaGlyLysGlyArgGlyGluGlyGlnArgMetMetAsn   235
                                                          |||
Db    1100 --------------------AGAATTGAAACGCAGAAACAACAACCCTCAAT   1129

Qy     236 ArgValAspArgMetGlnGln------ArgGluGlnGluGluGlnGluAspGluAspAsp   253
              :::      :::||||||      :::::::::::
Db    1130 GCCCCAGATGAGTCAATTAAAAAACTTCTGAGATGTTGCAAAGTAAGGCTTGCCATCG   1189

Qy     254 AspAspLeuGluAspValProArgArgArgSerArgAspGlyProGln---   272
              ||||||||||      :::::::::
Db    1190 AAAAGCCTGGAGGATGACAATGACCAACGCGGGATGGCAGAGGCTGAGTCTCAGGTC   1249

Qy     273 ------------------------------SerGluAlaGluHisGlnArgArgAsp   281
                                                          |||      |||
Db    1250 AGCCACTTGGAAGTGATTTTAGATCAGAAAGAGAAGAAGAAAACATACATCTTAGAGAGGAA   1309

Qy     282 LeuAlaArgLeuLysSerProArgLeuLysGluLeuLeuGlnLeuAsnAlaGluVal   301
              |||                                          :::::::::
Db    1310 TTGCACGGAGAAGCCAACTTCAGCCG---------------GAGCCAGCCAAGACG   1351

Qy     302 GlnSerLeuSerTyrGlnArgMetArgArgAspSerProLeuSerLysArgArgProLeu   321
              :::::::::                        :::::::
Db    1352 AAGGCTCTCCAGACTGTCATCGAAATGAAGGACACAAAAATCGCTTCA---------TTG   1402

Qy     322 AlaMetAsnAspGluAsp---GluSerAlaPheArgAlaMetGluAlaArgAlaLysLeu   340
              |||      ::::::::::::
Db    1403 GAACGAAACATAAGGGATCTTGACGATGAGTTAAAAGCCAATGGTGTGCTG   1462

Qy     341 AspGlnLysSerGlnLeuValLeuGlyLeuHisGlyPheGlyLeuSerArgAspAspGlu   360
              |||                                          :::
Db    1463 ------------------------ATCAAACAAATTGAGGTTTACAAAGTCACTCC   1519

Qy     361 AspGluAspGluAsnLeuIleLeuAspProSerGluAsnSerPheArgAlaArgProLeu   380
              |||                                          |||:::
Db    1520 AAGTTTATGAAGACCAAGATTGATCAGCTGAAGCAGGAACTTTCAAAGAAGAGTCGGAA   1579

Qy     381 ArgLeuSerSerGlyPheValGluValLysSerAsnAspGluLysSerAlaLeu   400
              :::      :::                                          :::
Db    1580 CTTGCCTTACAAACAAAAGTTGAAACCCTCAGCAAT   1618

Qy     401 AspArgIleLysTyrArgValAspArgValGluLysTyrLeuAlaProLysProMetGlu   420
              ||||||:::::::                        :::      :::
Db    1619 CAA---------------AATTCAGATTGCAAGCAACACATTGAAGTGCTCAAAGAG   1660

Qy     421 PheAsnProLysProGlyThrTyrPheAlaProLysIleProThrArgProArg   440
                                                          |||:::
Db    1661 TCACTTACT---------------------GCCAAAGAACAGAGGGGCTGCCATC---CTTCAGACTGAG   1705

Qy     461 ProSerThrGluTrpLysIleLysGluSerArgValLeuThrAsnLeuLysAsnAsn   480
              |||
Db     497 LeuGlyLysAsnThrLeuULys   497

Qy     481 ProSerLeuAlaAlaLeuPheMetAspAspLys------LeuGlyLysAsnThrLeuUlys   497
              |||                                          :::::::::
```

```
Db   1706 GTAGATGCGCTGAGATTACGACGTGGAAGAAAAAGAATCTTTCCTCAATAAAAAAACAAAA 1765
                                                              LeuValGlyLysSerHisGlnLysLeuProLeuAlaLeuAlaLeuGluLeu 55
Qy   36                                                       |||  |||  ||||||||||||||||||||
Db   632                                                       CTGACTGGCTCATCCATCATCACCACCACCAGGTCCCCTCATG---TTGAGGCAGGTA 688

Qy   498 GlyArgGlnMetLeuThrAspGluGlnLysGlyArgThrArgValLysThrIleArgAla 517
Db   1766 CAGCTACAGGACACCTCACA---GAAGAGAGGGG--------------------------- 1795

Qy   56                                                           MetGlyValGlnPheValAlaGluIleLysLysLysGlyGlnMetGluMetAlaLysGly 75
Db   689                                                           AGAGACGACAATGTTAGATCTTCAGGCCCCAGCTGAAAGAACTGCAGAGAGAATGAC 748

Qy   518 LeuProArgLeuPheGlyAlaProThrAlaLysAlaGluMetIleAspAlaAlaLysValPhe 537
Db   1796 ---------------------ACACTGGCCGGTGAAATTCGTGACATGAAAGATATG 1831

Qy   76                                                           AlaPheLysThrGlnLeuGluValLeuGluValLysLysValHisProAspGlnPheAspLysTyr 95
Db   749                                                           CTCCTCCGGAAAGAGCTAGACATCAAGCAGCAAATTGGATCTTCCATGAACAGTATT 808

Qy   538 GlnAspIleGluGlnArgIleProLeuProProPhePheGluPhePheLysProLysGlyArgHisThr 557
Db   1832 TTAGAGTGAAGGAAAGAAAAATCAATGTTCTTCAGAAAAGATTGAAAACTTGCAAGAA 1891

Qy   96                                                           LysLysLeuLysValLysAspAspLeuAlaAlaAspAlaValMetGlnAlnGlnAlaGluMetAla 115
Db   809                                                           AAGACTTTCTGAGCTCGAGCTTAAGAACGGAGAGAGTCTTGAGGAAAGAAGAGGAGCGCG 868

Qy   558 ArgLeuValArgTrpThrGlyAlaAsnGluLysGluLysGlyProGlyLeuGlySerArgPheIle 577
Db   1892 CAACTTAGG--------GATAAAGACAAGCAACTGACCAACCTGAAAGACACAGA------- 1936

Qy   116                                                           LysLeuLeuGlnProLysSerGlyAsnAlaPheIleAspMetLeuAsnGlyAsnGlyIleTyrIlePro 135
Db   869                                                           CGGATGTCTGTCCTCAAGGAGCAGATGAGGTTTCCCATGAAGAAAATCAGCACCTACAG 928

Qy   136                                                           IleGlySerIleArgGlyLeuGluGluAspAlaIleAlaIleArgThrGlnLeuAspArgMetGluAsn 155
Db   929                                                           TTG-----ACAATCCAGGCCCTTCAGATGACGTGCGGAACCCAGAGAGACTTCAACCAC 982

Qy   156                                                           ThrAspProSerGluGlnIleAlaLysAlaValMetAspLysPheGlnThrGlnIleIleLeu 175
Db   983                                                           CTCCTCCAGCAAGAGAGTGGCACCGGAGACGG--GAGCACTTCACCATGCGAGCTGACC 1039

Qy   578 LeuProSerLeuAspProThrMetProAlaLeuAsnThrAlaPheSerThr---------- 594
Db   1937 GTGAAGTCCTTGCAGACGGATTCCAGTAATACAGATACTGCCACTGGCCACGCTAGAGGAA 1996

Qy   176                                                           ProGlyLeuValAlaAlaAsnMetIleAlaGlyLysAsnProPheLysMetProGlnMet 195
Db   1040                                                           --------------GAGGAACTTTAGGCGGCTCCAAGCCGAG 1069

Qy   595                                                           GlnGlyArgAlaArgAspGlu 601
Db   1997                                                           GCTCTCGTCAGAGAAGGAGAAGGAGAATAATTGAGCGGTTGAAAGACAGCAGGAAAGAAGATGAT 2056

Qy   196                                                           ArgLysAlaGlnAlaAlaProSerValPheGlnGlnAlaLeuAlaGlnAlaGlnAlaMet 215
Db   1070                                                           CATGGACAGGCAGGCTAAGGAGCTGTTCCTTTTGAGGAAGACATTAGGGAAATGGAGCTG 1129

Qy   216                                                           LeuGlyLysAsnAlaProValAlaGlyLysValArgGlyValGluGlnArgMetMetAsn 235
Db   1130                                                           AGAATTGAAACGCAGAAACAAACCCTCAT 1159

Qy   602                                                           TrpAspThrMetPheLysIleProAsnAsnTrpAsnProGlyGluValGlyValGlyPheLys 621
Db   2057                                                           CGGGAAAGAGACTAGAAGAGAGATAGAATCCTTCCGAAAAGAGAACAAAGACCTGAAAGAGAAG 2116

Qy   236                                                           ArgValAspGlnArgMetGlnGln---------ArgGluLeuGlnGluGlnAlaAspAsp 253
Db   1160                                                           GCCCGAGATGAGTCAATTAAAAAACTTCTTGAGATGTTGCAAAGTAAGGCTTGCCATCC 1219

Qy   622 MetAsnSer 624
Db   2117 GTCAATGCT 2125

Qy   254                                                           AspAspLeuGluAspGluProValProArgAlaArgSerGluSerAspGlyProGln--- 272
Db   1220                                                           AAAAGCCTGGACGGATGACAATGAGCGACGCGGATGGGATGGCTGAGTCTCAGGTC 1279

Qy   273                                                           ------------------SerGluAlaGluHisGlnArgArgAsp 281
Db   1280                                                           AGCCACTTGGAAGTGATTTAGATCAGAAGAGAAGGAAAACATACATCTTAGAGAGAGAA 1339

Qy   282                                                           LeuAlaArgArgLeuLysSerSerProArgLeuLeuLeuGlnAsnAlaGluVal 301
Db   1340                                                           TTGCACCGGAAGAAGCCAACTTCAGCCG---------------GAGCCAGCCAAGACG 1381

Qy   302                                                           GlnSerLeuLeuSerTyrGlnArgMetArgAspSerProLeuSerLysArgArgProLeu 321
Db   1382                                                           AAGGCTCCAGACTGTTCATCGAAATGAAGGACACAAAAATGCTTCA------TTG 1432

Qy   322                                                           AlaMetAsnAspGluAsp--GluSerAlaPheArgAlaMetGluAlaArgAlaLysGlu 340
Db   1433                                                           GAACGAACATAAGGGATCTTGAGGATGAGATCCAGATGTTAAAAGCCAATGTGTGTCTG 1492

Qy   341                                                           AspGlnLysSerGlnLeuValLeuGlyValLeuHisGlyPheGlyGluSerAspAspAspGlu 360
Db   1493                                                           ------------------------AACACTGAG 1501

Qy   361                                                           AspGluGluAspGluAsnLeuIleAspProSerGluAsnSerPheArgArgAlaProLeu 380
Db   1502                                                           GACCGCGAAGAGAG--------------------------- 1516
```

RESULT 7

US-10-302-172-528

; Sequence 528, Application US/10302172
; Publication No. US20040053250A1
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Xue, Aidong J.
; APPLICANT: Drmanac, Radoje T.
; TITLE OF INVENTION: No. US20040053250A1e1 Arginine-rich Protein-like Nucleic Acids an
; TITLE OF INVENTION: Polypeptides
; FILE REFERENCE: 803_1CNCP
; CURRENT APPLICATION NUMBER: US/10/302,172
; CURRENT FILING DATE: 2002-11-21
; PRIOR APPLICATION NUMBER: US 10/225,251
; PRIOR FILING DATE: 2002-08-20
; PRIOR APPLICATION NUMBER: PCT US02/05095
; PRIOR FILING DATE: 2002-03-05
; PRIOR APPLICATION NUMBER: US 09/799,451
; PRIOR FILING DATE: 2001-03-05
; SOFTWARE: pt_FL_genes Version 2.0
; NUMBER OF SEQ ID NOS: 950
; SEQ ID NO 528
; LENGTH: 6151
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (263)..(3133)
US-10-302-172-528

Alignment Scores:

Pred. No.:		
Score:	169.50	Length:	6151
Percent Similarity:	36.28%	Matches:	118
Best Local Similarity:	18.94%	Conservative:	108
Query Match:	5.15%	Mismatches:	248
DB:	18	Indels:	149
		Gaps:	22

US-10-736-868-2 (1-643) x US-10-302-172-528 (1-6151)

```
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 27636
; LENGTH: 3603
; TYPE: DNA
; ORGANISM: Neurospora crassa
US-10-369-493-27636
```

```
Alignment Scores:
Pred. No.:                          Length:          3603
Score:                 1.36e-05     Matches:          165.00
Percent Similarity:    33.25%       Conservative:     120
Best Local Similarity: 19.15%       Mismatches:       259
Query Match:           5.01%        Indels:           310
DB:                    17           Gaps:             36
```

US-10-736-868-2 (1-643) x US-10-369-493-27636 (1-3603)

```
Qy   19 SerAlaGlnSerProSerThrSerAspAlaProGlyAlaLeuSerSer------ 35
            ::|||  |||||||||||:::||||||||
Db  203 ACGGGATCAAGCCCGTCTTCGTCTTTGACGGCGGCGGCCCCCGTCTCAAGCGCCGCCACCA 262

Qy   36 ------LeuValGlySerHisGlnLeuLeuProLeuAlaProSerMetGluAlaLeu 53
               |||  :::  :::  |||
Db  263 TCCAGGCTCGCCGCGAAGCAGAGGGCAGAGGGCAGGACGGCCACG-CGAACGGCCAGGG 321

Qy   54 GluLeuMetGlyValGlnPhe------------------------Val 61
             :::|||:::
Db  322 AAATTACTAGGCGGTGCAGATGCATCTATCGCAGAGGAGAGAACCGAGGGCCAAA 381

Qy   62 AspAlaLeuIleLeuLysGlyGlnMetGluMetAlaLysGlyAlaAlaPheLysThrGlnLeu 81
             |||
Db  382 AGGGCCAGAGCCACGAGACGCAATAAAACGGGCAGCAGCAGCAGCAGGAAGCAGAAGAA 441

Qy   82 GluValLeuGluLysValHisProAspArgLysGlyTyrLysLysLeuLysValAlaAsp 101
             |||  :::  ||||||
Db  442 GAA------------CAGGAGTACTGCCCGACATG----------GACAAAGTCGTTTACCCGGAC 486

Qy  102 AspLeuAlaAlaAspAlaValMetGlnAlaIleAlaGluMetAlaLysLeuGlnProLysSer 121
             :::::||:::  ||||||
Db  487 GAATTAGGCATGTCA-----AAACAAGAACGCCAAAAGTCTCCACAGCAA 540

Qy  122 GlyAsnAlaPheIleAspMetLeuAsnGlyAsn------------GlyIleProIleGlyIle 137
             :::|||::  :|||
Db  541 GACGCCTACCATCTTCCTGACCTCGACGCGGCATCGGCAAGCCGCGCCGATGAT 600

Qy  138 SerSerIleArgGlyLeuGluAspAlaIleArgArgMetGluAsnThrAsp 157
             :::||::||  |||
Db  601 CCCGCGCCATCATGTCGTCGAAGAACTCGAAGAGTACGTCAGGCAGTTTGAGAACGGGGAG 660

Qy  157 ------------------ 157
Db  661 GATATCAACTTGTACGATTTGCAGCAAGATCGACTTTGATGGCGAAGTTCTTCAAGTCTTTG 720

Qy  158 ---ProSerGluGlnIleAlaLysAlaValMetAspGlyPheGlnThrGlnIleLeuPro 176
            ||::::::::
Db  721 CCGCCGGTGATCGGTGATCGGTACAACATCTTGAATGCGCGAGTTGCGGAGCAGATTGAGAATG 780

Qy  177 GlyLeuValAlaAlaAsnMetIleAlaGlyLysAsnProPheLysMetProGlnMetArg 196
            ||||||  :::
Db  781 GGGTTGAGTAAAGAACCAGTTGGACGTCATGTTCCCTACCCGGATTGGATTTCAGCCGGTTT 840

Qy  197 LysAlaGlnAlaAlaProSerSerValPheGlnGlnAlaLeuAlaGlnAlaMet--- 215
             :::
Db  841 CAGATTGAGAGAGTCAAGGAG------------AGGAATCACTTGACCAGAGACTGATGTAC 891

Qy  215 215
Db  892 GAGATGGGCATGACGGGATTGATCTGACCTGGCAGTCAACCGCGACGAGAGTCGCGGGC 951

Qy  216 ------LeuGlyLysAsnAlaProValAlaGlyValGlyArgGly--- 227
            |||:::  |||
Db  952 GATCAGAATAGGAGTATATCCTGGTCAAGATGATATTGCCGCCGGAATCGAAGGTGCCGAAGGCGGATATGCGCTG 1011
```

```
Qy  381 ArgLeuSerSerGlyPheValGluLysLeuLysSerAsnAspGluLysSerAlaLeu 400
             :::|||  ||||||
Db 1517 ------------------ATCAAACAAATTGAGGTTACAAAGTCACTCC 1549

Qy  401 AspArgIleLysTyrArgValAlaAspValGluLysTyrLeuAlaProLysProMetGlu 420
            :::|||  :::::::|||  |||
Db 1550 AAGTTTATGAAGACCAAGATTGATCAGCTGAAGCAGAACTTCAAAGAAGAGTCGGAA 1609

Qy  421 PheAsnProLysProGlnProGlyTyrPheAlaProArgLysIleProThrArgProArg 440
            :::|||
Db 1610 CTT------------------CTTGCCTTACAAACAAGCTTGAAACCCTCAGCAAT 1648

Qy  441 LysMetLeuProLeuIleLysSerAspProLysValGlnGluIleLeuArgArgHis 460
            :::
Db 1649 CAA------------------AATTCAGATTGCAGCAACACATTGAAGTGCTCAAAGAG 1690

Qy  461 ProSerThrGluTrpLysIleLysGluSerArgValLeuThrAsnLeuLysAsnAsn 480
             |||||||
Db 1691 TCACTTACT------------GCCAAAGAACAGAGGGCTGCCATC---CTTCAGACTGAG 1735

Qy  481 ProSerLeuAlaAlaLeuPheMetAspAspLys------LeuGluAsnThrLeuLys 497
             :::::::||  :::::
Db 1736 GTAGATGCCGTCGAGATTACGACTGAGAAGAAAAAATCTTCCTCAATAAAAAAACAAA 1795

Qy  498 GlyArgGlnMetLeuThrArgAspGluLysGlyArgThrArgValLysThrIleArgAla 517
            ||||||:::||||||
Db 1796 CAGCTACAGGAGACCTCACA--GAAGAGAGGGG 1825

Qy  518 LeuProArgLeuPheGlyAlaThrAlaLysAlaGluMetIleAspAlaLysAlaValPhe 537
             |||
Db 1826 ------------------ACACTGGCGCCGGTGAAATTCGTGACATGAAAGATATG 1861

Qy  538 GlnAspIleGluGlnProIleProLeuPhePheGluProLysGlyArgHisThr 557
             :::::: :::
Db 1862 TTAGAAGTGAAGGAAAGAAAAATTCAATGTTCTTCAGAAAAAGATTGAAAACTTGCAAGAA 1921

Qy  558 ArgLeuArgTrpThrGlyThrAlaAsnGluLeuLysGluIleProGlyLeuGlySerArgPheIle 577
            :::||||||:::::::
Db 1922 CAACTTAGG------------GATAAGACAGCAACTGACCAACCTGAAAGACAGA 1966

Qy  578 LeuProSerLeuAspProThrMetProAlaLeuAsnThrAlaPheSerThr 594
            :::  ||||||
Db 1967 GTGAAGTCCTTGCAGACGGATTCCAGTAATACAGATATCGCACTGGCGACGCTAGGAGGAA 2026

Qy  595 ------------------GlnGlyArgAlaArgAspGlu 601
                  :::  |||||||:::::
Db 2027 GCTCTGTCAGAAGAGGAGGAGAATAATTGAGCCGCTTGAAAGAACAGCGAGAAGAGATGAT 2086

Qy  602 TrpAspThrMetPheLysIleProAsnTrpAsnProGlyAspGluValGlyValGlyPheLys 621
            :::  :::|||
Db 2087 CGGGAAGAGCTAGAAGAGATAGAATCCTTCCGAAAAGAGAACAAAGACCTGAAAGAGAAG 2146

Qy  622 MetAsnSer 624
            :::|||
Db 2147 GTCAATGCT 2155
```

```
Qy    228 ---------------GluGluGlnArgMetMetAsnArgValAspGlnArgMet 241
                         :::::|||||        |||  |||
Db   1012 GGAGTCTGAGTAGTACAGAGAAGGATAAAGGACAGATGACCAATTCCAATTGAT------GTG 1065

Qy    242 GlnGlnArgGluLeuGln----------GluGluAspGluAspAspAspAsp 255
              ::: ||| |||               |||||:|||: ::::||||||
Db   1066 GACGAACTCGAGTTTCAGTTTCAGGGCAAGGAGGAGGAGGAGGAAGAGGAT 1125

Qy    256 LeuGluAspGluValPro----------ArgArgArgSer 266
          :::|||| ::::|||||             |||||||:::
Db   1126 GTCGAGGTTCGAGGACGTGCCAATTGAGCGGGTTGAATCGGTTGCCAAGCTGACTAGGGCT 1185

Qy    267 SerAspGlyGluProGlnSerGluAlaGluHisGlnArgArgAspLeuAlaArgArgLeu 286
          ::: |||||:::            ::|||||:
Db   1186 GCTCGGTTACAGGCCAGGGGGGGACTGCGAAT----------TCGAGGAGGCAG 1227

Qy    287 LysSerSerProArgLeuLysGluLeuLeuGlnAsnAlaGluValGlnSerLeuLeuSer 306
          ||||| ||| :::                     :::::: :::
Db   1228 TTCTACGAGCCAGACAGAGTGAAGAGGCCGAGGACGAGGAC--------- 1269

Qy    307 TyrGlnArgMetArgAspSerProLeuSerArgArgProLeuAlaMetAsnAspGlu 326
          |||||                                   :::::
Db   1270 ----------------------------------------------GAAGAT 1275

Qy    327 AspGluSerAlaPhe 334
          |||||||:::
Db   1276 GACGAGGCCTTGTTTGTTGACAACCGACAAGATGCCGTCTATGGAGGAGGAATGCGGTC 1335

Qy    335 GluAlaArgAlaLeuLysAspGlnLysLysSerGlnLeuValLeuGlyLeu----His 351
          ||||||   ::: |||                                       |||
Db   1336 GAAGCTGATGAGGAAGAAGACCTCAACAAGGCTATTGCCATGTCTCTTCAGAATCAGCAT 1395

Qy    352 GlyPheGlyLeuSerAspAspAspAspGlu----AsnLeuIleAspPro 370
          :::||||| ||| ||| |||:::            ::: :::|||||
Db   1396 GGTGTTGGTCTATCAGAAGACGAGGATATGGAGGTTCGAAGAGGTTACTCCG 1455

Qy    371 SerGluAsnSerPheArgArgAla----------ProLeuArgLeuSerGlyPhe 386
          ||| :::             |||::          |||||||||||:::
Db   1456 CCAGAGTGGGACCAGAAAGCCGTGGAAGCGCCCAAGCCCATTGTAGTACCAGTGTCGC 1515

Qy    387 ValGluLysLeuLysSerAsnAspGluLeuLysSerAlaLeuAspArgIleLysTyrArg 406
          ::: |||||:::                 ::|||||          :::
Db   1516 ATGGTGCGACACATTGTCAACAACAGGGGAGTGCTGCCGTCCGTCCAAGGCGAAGGGAGT 1575

Qy    407 ValGluAspValGlu 411
          |||||||||:::
Db   1576 ACCGACAGCAGTGACAGCGCGAGGGGGATATTCGAGCAGCCCATGGCCCCTGCCAGGAGGAAG 1635

Qy    412 -----LysTyrLeuLeuAlaProLysProMetGlnPheAsn 422
               :::|||||                   |||:::
Db   1636 CAAGCAAAGGCTCCAGCGCGCCAAAGCCCAGTGGTCGAAAAACCGCCCTGCTCCTACCAACACT 1695

Qy    422 ProLysPro---GlnPro 427
          ||||||     |||:::
Db   1696 CCCTTTAAAGGTCCCCTTCATTTGAGAAGCTGAACTGGAAGCAACTGCTGGGGCAAC 1755

Qy    428 GlyTyrPheAlaProArg 437
          :::|||::
Db   1756 CGACCGAAAGCAAAATCTCAGAAGAGAAAAGGTGTCTGAACCCAAGCCAGTTGAAACT 1815

Qy    438 ArgProArgLysMetLeuProLeuIleGlySerAspProLysValGlnGluGluIle 457
          :::|||||   :::      ::::::    :::  ||
Db   1816 GCTCCTCGTCCATGGTCCACTACCGCCATGGTCTAGCGAATAACGACGAAGACACTGTA 1875

Qy    458 ArgArgHisProSerThrGluTrpLysIleAlaLysGluSer 471
          ||:|::
Db   1876 GGTGAGATGGAACCTCAGTCTGATGACGATGCTGGCCGTTTTGAGAAAGAGCCGAAGGAT 1935

Qy    471 ------------------------------ 471
          :::
Db   1936 CGGAAGCAGCGAGAACAGGACAGGCAGCAGCAGCAGCGAGGATGAAAACTTGGCTCCAGAA 1995
```

```
Db   1996 GAGCAGCGGCCAGAATTCGAAAGCAACGAGAGGATGAGGTCATCGATGATTGAGTCATCCGAT 2055

Qy    472 --------ArgValLeuThrAsn 476
                  :::|||||||
Db   2056 GAAAGTGATGGTGATTTGGAGATCCTCGACGTCCTCCCATCGCCCAAGAGCTCACTCAG 2115

Qy    477 LeuLysAsnAsnProSerLeuAlaAlaLeuPheMetCysAspAspLysGluLeu------GluAsn 494
          :::        |||||||:::|||                  |||
Db   2116 ATGCCG---GGGCCTAGCTTGCCCTTCTTTG-----GACGACGAAGTGTCGCAAGAGCCA 2166

Qy    495 ThrLeuLysGlyArgGlnMetLeuThr---AspGluGlnLysGlyLysArgThrArgValLys 513
          ||| |||||:::              :::       |||
Db   2167 GTTGAAACCAGCCGTCAATCACCCTCCCTAATGAGGAAGAGGCGACGACATTGTGTTT 2226

Qy    514 ThrIleArgAlaLeuProArg---------LeuPheGlyAlaProThrAlaLys 528
          |||||   :::                  ||||||
Db   2227 GAAGACGTAACCCTTCCGTCTCAAACCATGACCCTCCTTCGGACCACCCGCAACTCAA 2286

Qy    529 AlaGlu----------MetIleAspAlaLysValPheGlnAspIleGluGluArgProIle 545
          ||||||          |||:::||:
Db   2287 GCCGAAGAAGACGCCCCAGATCGAAGCAGACCTCTTCAAAGACGTC-------------CTT 2334

Qy    546 ProProLeuPhePheGluProLysProAlaProHisThrArgLeuArgHisThrGlyAlaAsn 565
          |||||:::                                        |||||
Db   2335 CCTCCA----------------------------------------------CCTGCCCAA 2349

Qy    566 GluLysGluIleProGlyLeuGly----------SerArgPheIleLeuPro 579
          :::::::|||                       :::|||||:::|||||
Db   2350 CAGGAACAGTTGAAGTTTGGCGAGGTGGGTTCCAGATCGACATCGTTCCC 2409

Qy    580 SerLeuAspProThrMetProAlaLeuAsn---ThrAlaPheSerThrGlnArgGlyAlaAr 599
          ||| :::|||                   :::  ||||::: :::
Db   2410 ------GACCCCGACGCGCCCCCTCTCCGAATTCGAAGACGACTTCAGCGACCCCGAA 2463

Qy    599 gAspGluTrpAspThrMetPheLysIleProAlaAsnTrpAsnProGlyAspGluValGl 619
          :::||::                                  |||||||||
Db   2464 GAACAAGAACTCCTCG-----------------CTTCCCTCGCCCCAAGAAGCCGAAGAACACGCCCGC 2514

Qy    619 yPheLysMetAsnSerLysThrLysArgPhe 629
          :::|||||||::
Db   2515 TTCGCCACTCAACTCAACCACAAAACCTTTC 2545
```

RESULT 9
US-10-747-065-1

; Sequence 1, Application US/10747065
; Publication No. US20040171810A1
; GENERAL INFORMATION:
;   APPLICANT: Ohtsuka, Toshihisa
;   APPLICANT: Rikitsu, Etsuko
;   APPLICANT: Inoue, Marie
;   APPLICANT: Inoue, Eiji
;   TITLE OF INVENTION: Presynaptic Protein CAST
;   FILE REFERENCE: 2144.0160000
;   CURRENT APPLICATION NUMBER: US/10/747,065
;   CURRENT FILING DATE: 2003-12-30
;   PRIOR APPLICATION NUMBER: JP 2002-063186
;   PRIOR FILING DATE: 2002-03-08
;   PRIOR APPLICATION NUMBER: PCT/JP03/02718
;   PRIOR FILING DATE: 2003-03-07
;   NUMBER OF SEQ ID NOS: 8
;   SOFTWARE: PatentIn Ver. 2.1
;
; SEQ ID NO 1
;   LENGTH: 2932
;   TYPE: DNA
;   ORGANISM: Rattus norvegicus
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (59)..(2929)
US-10-747-065-1

Alignment Scores:
Pred. No.:   1.45e-05   Length: 2932

```
Score:                    163.50    Matches:           118
Percent Similarity:        36.17%   Conservative:      107
Best Local Similarity:     18.97%   Mismatches:        250
Query Match:                4.97%   Indels:            147
DB:                                 Gaps:               22

US-10-736-868-2 (1-643)  x  US-10-747-065-1 (1-2932)
```

```
Qy      36 LeuValGlyLySerHisGlnLysLeuProLeuAlaProSerMetGluAlaLeuGlu  55
              |||    |||||  ||||||
Db     428 CTGGTGGCTCATCCACCACCACCAGGTGCCCTCCATG---TTGAGGCAGGTA 484

Qy      56 MetGlyValGlnPheValAspAlaLeuIleLysLysGlyGlnMetGluMetAlaLysGly  75
                           ::   |||||
Db     485 AGAGACAGCACGATGTTGGATCTTCAAGCCAGCTCAAAGAACTGCAGAGAGAAATGAC 544

Qy      76 AlaPheLysThrGlnLeuGluValLeuGluLysValHisProAspGlnPheAspLysTyr  95
              :: :: ::|||||| ||||:::  ::
Db     545 CTCCTTCGGCAAGGAGCTTGACATTAAGGACAGCAACAACTGGGGTCTTCATGAATAGCATC 604

Qy      96 LysLysLeuLysValAspAspLeuAlaLeuAlaValMetGlnAlaGlnGluMetAla  115
              |||        ::  ||  ||:::::::::  |||
Db     605 AAGACTTTCTGAGTCCTGAGCTTAAGAGCAGAGTCTTGAGGAAAGAGGAGGCTGCT 664

Qy     116 LysLeuGlnProLysSerGlyAsnAlaPheIleAspMetLeuAsnGlyAsnGlyIlePro  135
              :::    ||
Db     665 CGCATGTCTGTCCTCAAGGAGCAGATGAGGGTTTCTCACGAAGAAAACCAGCACCTGCAG 724

Qy     136 IleGlySerSerIleArgGlyLeuGluAspAlaIleArgThrGlnAspMetGluAsn  155
              ::|||:::||  ||  |||:::||| 
Db     725 TTG-----ACCATCCAGGCCCTTCAGGATGAACTGCGAGACCCCAGAGACCTCAACCAC 778

Qy     156 ThrAspProSerGluGlnIleAlaLysAlaValMetAspLysPheGlnThrGlnIleLeu  175
              |||             :: ||
Db     779 CTCCTGCAGCAGGAGAGTGGCAACCGAGGAGCA---GAGCATTCACCATGCAGCTGACG 835

Qy     176 ProGlyLeuValAlaAsnMetIleAlaGlyLysAsnProPheLysMetProGlnGlnMet  195
              :: |||
Db     836 ---GAGGAGAACTTCCGCAGGCTCCCAAGCCGAA 865

Qy     196 ArgLysAlaGlnAlaAlaProSerSerValPheGlnIleAlaLeuAlaLeuAlaArgAlaMet  215
              ||||||||||  || 
Db     866 CACCGACAGGCAGGCCAAGGAGCTGTTCCTTCGCGGAACATTGGAAGAAATGGAGCTA 925

Qy     216 LeuGlyLysAsnAlaProValAlaGlyGlyIleAlaGlyIleAlaGlyLysLeuGlnArgMetMetAsn  235
              ||||||  :: :::|||
Db     926 AGGATTGAAACACAGAAACAGACTCTCAAT 955

Qy     236 ArgValAspGlnArgMetGln------ArgGluLeuGlnGluGluGlnGlnAlaGluVal  253
              ||| :::::::::                  :: :::
Db     956 GCCCGGACGACGGCTCCATTAAAAAGCTCCTGCAGGTAAGGGCTTGCCATCG 1015

Qy     254 AspAspLeuGluAspValProAspAspArgArgSerAspArgGlyGlyGln---  272
              ||||||| ::::: ||:::::
Db    1016 AAAAGCCTAGAGGACGACCAACGAGCGCCACGCGCGGATGGCCCGAGGCTCAGTCTTCAGTC 1075

Qy     273 ------------SerGluAlaGluHisGlnArgArgArgAsp  281
                          |||||:::
Db    1076 AGCCACTTGGAACTGATTTTAGACCGAAGGAGAAGGAGAAACATCCACCTGAGAGAGAA 1135

Qy     282 LeuAlaArgArgLeuLysSerSerProArgLeuLysSerGluLeuLeuGlnLeuLeuAsnAlaGluVal  301
              |||  :::  ||| |||:::       ::: ||||:::
Db    1136 TTGCACCGAAGAGAGCCAACTTCAGCCG-----------GAGCCAGCCAAGACG 1177

Qy     302 GlnSerLeuLeuSerTyrGlnArgMetArgAspSerProLeuSerLysArgArgProLeu  321
              |||:::  |||:::::   :::::
Db    1178 AAGGCGCTCCAGACTGTCATCGAAATGAAGACACAAAAATTCTTCA--------CTG 1228

Qy     322 AlaMetAsnAspArgGluAspGluSerAlaPheArgAlaMetGluAlaArgAlaLysLeuAsp  341
              |||||  ||:::::  :::
Db    1229 GAGCCGGAACATCCGGGAC------------------CTCCGAG 1252

Qy     342 GlnLysSerGlnLeuValLeuGlyHisGlyLeuLeuGlyGlySerAspAspGluAsp  361
```

```
Db    1253 GATGAGATCCAGATG------TTGAAAGCCAATGGTGTGCTCAACACAGAGGACCGAGAGGAG 1309

Qy     362 GluGluAspGlnAsnLeuIleLeuAspProSerGluAsnSerPheArgArgAlaProLeuArg 381
              :::                ::: |||       ::: |||
Db    1310 GAGATCAAACAGATCAGGTGTACAAAAGCCACTCCAAGTTTATGAAGACC---------- 1360

Qy     382 LeuSerSerGlyPheValGluLysLysSerAsnAspGluLeuLysSerAlaLeuLysAsp 401
              
Db    1361 ------------------------------AAGAATGACCAGCTGAAA------------ 1378

Qy     402 ArgIleLysTyrArgValAspAspValGluLeuTyrLeuAlaProLysProMetGluPhe 421
              
Db    1379 ----------------------------------------CAGGAACTT---------- 1387

Qy     422 AsnProLysProGlnProGlyTyrPheAlaProArgLysProThrArgProArgLys 441
              :: ::: |||:::                 :::
Db    1388 TCCAAGAGGAGTCAGAACTTCTGCCTTACAACAAAGCCTGAAACCCTTAGCAATCAG 1447

Qy     442 MetLeuProLeuLeuPheGlySerAspProLysValGlnLeuGluGlnIleLeuArgHisPro 461
              :::                 ||||||
Db    1448 ------------------AATTCAGATTGCAAGCAACACATTGAAGTGCTTAAAGAGTCA 1489

Qy     462 SerThrGluTrpLysIleAlaLysGluSerArgValLeuThrAsnLeuLysAsnAsnPro 481
              |||                      LeuGluAsnThrLeuLysGly 498
Db    1490 CTTACT-----------------GCCAAAGAACAGAGGGCTGCCATC---CTTCAGACTGAGGTA 1534

Qy     482 SerLeuAlaAlaLeuPheMetAspAspLys 
              |||          :::::::::::|||
Db    1535 GATGCCACTGAGATTACGGCTGGAAGAGAAGAATCTTTTCTCAATAAGAAGAAAACAAAACAG 1594

Qy     499 ArgGlnMetLeuThrAspGluGlnLysGlyIleLeuProGlyLeuLysSerArgPheIleLeu 518
              ||| |||                            ArgValLysThrIleArgAlaLeu 518
Db    1595 CTCCAAGACCTCACT--GAAGAGAGGG 1621

Qy     519 ProArgLeuPheGlyAlaProThrAlaLysAlaGluMetIleAspAlaLysValPheGln 538
              |||                   ||AlaLysValPheGln 538
Db    1622 ------------ACCCTAGCTGGAGATCCCGTGATATGAAAGATATGTTA 1660

Qy     539 AspIleGluArgProIleProProLeuPhePheGlyProLysGlyArgHisThrArg 558
              ::::::|||||| ::: |||
Db    1661 GAAGTAAAGGAAGAAAAATCAATGTTCTTCAGAAAAAATTGAAAACTTGCAAGAACAA 1720

Qy     559 LeuArgTrpThrThrGlyAlaAsnGluLysLysGluIleLeuProGlyLeuGlySerArgPheIleLeu 578
              ::: |||               :::|||
Db    1721 CTTTAGG-----------GATAAGGACAAACAACTGACCAACCTGAAAGACAGA------GTG 1765

Qy     579 ProSerLeuAspProThrMetProAlaLeuAsnThrAlaPheSerThr------ 594
              ::||||                       ::|||
Db    1766 AAGTCCCTGCGAGACGGACTCCAGCAACACTGACACTGCTCTGGAGGAGGCC 1825

Qy     595 ------------------------GlnGlyArgAlaArgAspGluTrp 602
                                      ::: ||||||
Db    1826 TTGTCGGAAAAGGAGAGAATAATAGAGCCGTTGAAAGAGCAGAGGAGGAGAGGATGATCGG 1885

Qy     603 AspThrMetPheLysIleProAlaAsnTrpAsnProGlyLysValGlyPheLysMet 622
              :::::::||                     ::|||
Db    1886 GAAAGACTAGAAGAGATAGAATCCTTTCGAAAGGAGAACAAAGACCTCAAAGAGGAGGTC 1945

Qy     623 AsnSer 624
              |||:::
Db    1946 AATGCT 1951
```

RESULT 10
US-10-084-817-158
; Sequence 158, Application US/10084817
; Publication No. US2003011909A1
; GENERAL INFORMATION:
; APPLICANT: Susan Stuart
; APPLICANT: Jed G. Nuchtern
; APPLICANT: Sharon E. Plon
; APPLICANT: Jason M. Shohet
; TITLE OF INVENTION: GENES REGULATED BY MYCN ACTIVATION

```
;  FILE REFERENCE: PA-0046 US
;  CURRENT APPLICATION NUMBER: US/10/084,817
;  CURRENT FILING DATE: 2002-02-25
;  PRIOR APPLICATION NUMBER: 60/270,784
;  PRIOR FILING DATE: 2001-02-23
;  SOFTWARE: PERL Program
;  NUMBER OF SEQ ID NOS: 365
;  SEQ ID NO 158
;  LENGTH: 6354
;  TYPE: DNA
;  ORGANISM: Homo sapiens
;  FEATURE:
;  NAME/KEY: misc feature
;  OTHER INFORMATION: Incyte ID No. US2003011909A1 1393778CB1
US-10-084-817-158

Alignment Scores:
  Pred. No.:              4.12e-05        Length:        6354
  Score:                  163.50          Matches:       170
  Percent Similarity:     30.60%          Conservative:  91
  Best Local Similarity:  19.93%          Mismatches:    249
  Query Match:            4.97%           Indels:        345
  DB:                                     Gaps:          35

US-10-736-868-2 (1-643) x US-10-084-817-158 (1-6354)

Qy    12 GlyPheCysIleAlaProLeuSer------AlaGlnSerProSerThrSerAsp------    27
         ||||||          |||         :: ::
Db  4068 GGCTTCTCAGCCAGTCCAGACAAGTCCAGCAAGCTTCACCAAGGACTTCTCCGGCCGTGG  4127

Qy    28 AlaProGly-----AlaLeuLeuSerLeuValGlyLysSerHisGlnLysLeuPro        45
         :::||                ||||||          :::::
Db  4128 AGTCCCAGCTGCGAGGACACTCAGGAGCTG-CTGCAGGAGGAACCGGCAGAAGCTGAGC   4186

Qy    46 LeuAlaProSerMetGluAlaLeuGluGlu----                            54
         ||||||            ::||
Db  4187 CTGGAGCACCAAGCTCAAGCAGCAGGAGGAAGAATTCCTTCCGGGAGCAGCTGGAG     4246

Qy    55 -----------------LeuMetGlyValGlnPhe                            60
                            |||||
Db  4247 GAGGAGGAGGAGGCCCAAGCACAACCTGGAGAAGCAGATCGCCACCTCCATGCCCAGGTG  4306

Qy    61 ValAspAlaLeuIleLysLysGlyGlnMetGluMetAlaLysProLysThrGln         80
         |||             :: :||||||
Db  4307 GCCGAC------ATGAAAAAGAGATGGAGGAACCAGCAGTGTGTGGGGTGCCTGGAAACTGCT  4357

Qy    81 LeuGluValLeuGluLysValHisProAsp---                             90
         |||||                    :::|||
Db  4358 GAGGAGGTGAAGAGGAAGCTCCAGAGGACCCTGAGCGCGGCCACGAGGAG           4417

Qy    91 ------GlnPheAspLysTyrLysLysLys-----------ValAsp               101
                 |||||                  :::||||:||
Db  4418 AAGGTGGCCGCGCTACGACAAGCTGGAGAAGACCAGCAGGAGCTGGAC             4477

Qy   102 AspLeuAlaAlaAspAlaValMetGlnAlaValMetGlnAlaLeuGlnProLysSer     121
         |||||||                   :::::
Db  4478 GACCTGCTGGTGACCGAGAAGCCGTCAGCGGCCCTGAGCGCGGCCCAACCTGAGAGCAG  4537

Qy   122 GlyAsnAlaPheIleAspMetLeuAsnGlyAsnGlyIleProIleGly---------    137
         |||                          :::
Db  4538 AAGAAG------TTTGACCAGCTGCCTGGCGGGAGGAAGAACCATCTCTGCCAAGTATGCA  4591

Qy   138 ----------SerSer                                              139
                     :::::
Db  4592 GAGGAGGCGCACCGGGCTGAGCGCGGCCCAGAGGAGCAAGGCTCTGCTG            4651

Qy   140 IleArgGlyLeuGluAspAlaIleAlaArgThrGlnAlaIleArgAspMetGluAsnThrArgAspProSer  159
         ||||||||       :::
Db  4652 GCCCGGGACGCGCCATGGAGGAGGACGGCCTCAAC                          4705

Qy   160 GluGlnIleAlaLeuAlaValMetAspLysPheGlnThrGlnIleLeuProGlyLeuVal  179
         :::||||:::||||                       :::
Db  4706 AAGCAGTTCCGCACGGAG-                                           4726

Qy   180 AlaAsnMetIleAlaGlyLysAsnProPhe---LysMetProGlnGlnMetArgLysAla  198
         :::       |||        ::::::|||
Db  4727 GAGGACCTTATGAGCTCCAAGAAGATGATGGGCAAGAGTGTGTCCAGACTGGAGAAGTC-  4785

Qy   199 GlnAlaAlaProSerSerValPheGlnAlaLeuAlaLeuAlaGlnAlaMetGluLeuGlyLys  218
                                                            -AGAGCAGGTGGAGGA
Db  4786 CAAGCGGGCCCT                                                  4815

Qy   219 Asn----------AlaProValAlaLeuAlaGlyTyrGlyTyrGluGluGlnLysArgMetMetAsn  235
         :::          |||||||||||
Db  4816 GATGAAGACAGCTGGAAGAGCTGGAGGACGAGGCTGCAGCGCACCGAAGATGCCAAGCT  4875

Qy   236 -ArgValAspGlnArgMetGln----------------GlnArgGluGlnGluGl      249
          ::|||||||:::|||
Db  4876 GCGGGTTGGAGGTCAACCTGCCAGGCCATGAAGGCCCAGTTGAGGCCCTGGAGGGGCCG  4935

Qy   249 uAspGluAspAspAspLeuGluLeuAspSerValProArgArgSerAspGl          269
         |||||     :: :||||||              |||||
Db  4936 GGACGAGCAGCGAGCGGAGGAAGAAGCAGCTGGTCAGCAGGGTGCGGGAGGATGAGGC  4995

Qy   269 yGluProGlnSerGlnSerGlnHisGlnArgArgAspArgAspArgLeuGlyLeuLysSerSe  289
         |||             :: :::                    :::||
Db  4996 AGAGCTGGAGCAGCAGGAAGCAGCGCTCGATGGACGCCGCCCCGGAAGAAGCTGA      5055

Qy   289 rProArgLeuLysGluLeu------                                     295
         ||||||:::||
Db  5056 GATGGAGACCTGAGGACCTGAGCGGCCGCACATCCGCCAACAAGAACCGGACGACGAGC  5115

Qy   296 ------LeuGlnAsnAlaGluValGlnSerHisLeuLeuGlnSerTyrGlnArgMetArgAs  312
               ::: :::       ||||||
Db  5116 CATCAAACAGCTGCGGAAGCTGCAGCGGCCCAGATGAAGGAGCTGCGGAGCTGGATGA   5175

Qy   312 pSerProLeuSerLysArgArgProLeuAlaMetAsnAspGluSerAlaPheAr        332
         |||::                     :::
Db  5176 CACCCGCCCTTCGTGAGGAGGATCCTGGCCCAAAGGAGAAGAAGGTGAA            5235

Qy   332 gAlaMetGluAlaArgAlaLysLeuAspLysArgGlnLeuValLeuGlyLeuHisGl     352
         |||                       :::||||||||
Db  5236 GAGATGATCCAGTTGCAGGAGGAACTGGCAGC                             5280

Qy   352 yPheGlyLeuSerAspAspAspAspGluGluAspGluAsnLeuIleLeuAspProSerGl  372
         |||::                     ||||||||
Db  5281 CGCGGAGCGGTGCCAAGCGCCAGGCCACGAGGAGATGATGCTGGCTGCGACGAGATCGC  5340

Qy   372 uAsnSerPheArgArgAlaProLeuArgLeuSer------                      383
         |||:::                ||||||
Db  5341 CAACAGCAGAAGCGGCCAAAGAGCCCGTTAGGAGGAAGCGGCCCGTCTGGAGGCCCCAT  5400

Qy   384 -SerGlyPheValGluLysLeuLysSer-----------AsnAspGl              395
          :::                :: :::||||:||
Db  5401 CGCCCAGCTGGAGGAGGACTGGAGGAGGAGCTGCATCAACGACCG                5460

Qy   395 uLeuLysSerAla-----                                            399
         |||::
Db  5461 GCTGAAGAAGCCAACCTGACAGATCGACACACCGACCTGGAGCGCAG              5520

Qy   400 --------------LeuAspArg-----IleLeuy                           404
                       ||||:||       :::||||y
Db  5521 CCAGCCCCAGGAAGAACAGAGAATGCTCGGCGGGAGGAAGAACAACAGAGCTTAA      5580

Qy   404 sTyrArgValAspGluGlu-----LysTyrLeuAla-ProLysPro               418
         s::::                   |||::       ||||||
Db  5581 GGTCAAGCTGCAGGAGGTGGAGGGCACTGCTCAAGTCAAGGGCTCCATCACCGC       5640

Qy   419 --------------MetGluPheAsnProLysPro                           425
                       ||||||:::
Db  5641 CCTCGAGGCCAAGATTGCCAGCTGGAGGAGGAGCAGCTGGACAACGAGCAGGCGCCA    5700

Qy   426 --GlnProGlyTyr                                                429
           |||||||
Db  5701 GGCAGCCTGCAAACAGTGCGTCCGGACACGGAGAAGGAAGGATGTGCTGCTGCAGGT    5760
```

```
;   NAME/KEY: misc_feature
;   LOCATION: (162)..(162)
;   OTHER INFORMATION: n equals a,t,g, or c
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (4022)..(4022)
;   OTHER INFORMATION: n equals a,t,g, or c
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (4039)..(4039)
;   OTHER INFORMATION: n equals a,t,g, or c
US-10-264-049-543

Alignment Scores:
Pred. No.:              2.51e-05           Length:              4041
Score:                  163.00             Matches:             177
Percent Similarity:     30.23%             Conservative:        89
Best Local Similarity:  20.11%             Mismatches:          262
Query Match:            4.95%              Indels:              354
DB:                     17                 Gaps:                35
```

US-10-736-868-2 (1-643) x US-10-264-049-543 (1-4041)

```
Qy      12  GlyPheCysIleAlaProLeuSer------------AlaGlnSerProThrSerAsp------  27
                ||||:::      ||:||                :::        |||:|||||
Db     582  GGCTTCTCAGCCAGTCCGACAGCAGTCCAGCAAGCTCACCAAGGACTTCTCCGGCGCTGG  641

Qy      28  AlaProGly-----------AlaLeuLeuSerSerLeuValGlyLysSerHisGlnLysLeuPro  45
                :::|||            |||||||||||           :::::||   |||||:||
Db     642  AGTCCCAGCTGCAGCACACTCAGGAGCTG-CTGCAGGAGGAGAACCGGCAGAAGCTGAGC  700

Qy      46  LeuAlaProSerMetGluAlaLeuGlu--------------------------------  54
                 ||:|||:  ::::  ||:||:||
Db     701  CTGGAGCACCAAGCTCAAGCAGTGGAGGACGAGAGAATTCCTTCGGGAGCAGCTGGAG  760

Qy      55  ------------------------------LeuMetGlyValGlnPhe----------  60
                                             :::  |||
Db     761  GAGGAGGAGGAGGCCAAGCACAACCTGAGAAGCAGATGCCACCCTCCATGCCCAGGTG  820

Qy      61  ValAspAlaLeuIleLysLysGlnMetGluMetAlaLysPheLysThrGln----------  80
                 :::||||||||||  |||  :::||      ::::|||||
Db     821  GCCGAC-----------ATGAAAAGAATGGAGGACAGTGTGGGGTGCCTGGAAACTGCT  871

Qy      81  LeuGluValLeuGluLysValHisProAsp------------------------------  90
                |||:::||||  ::::||||||
Db     872  GAGGAGGTGAAGAGGAAGTCCAGAAGGACCTGGAGGGCCTGAGCCAGCCGCACGAGGAG  931

Qy      91  ------------GlnPheAspPheTyrLysLysLeuLeuLys------ValAsp------  101
                            :::||||:::|||||||            |||:||
Db     932  AAGGTGGCCGCCTACGACAAGCTGGAGAAGAACCAAGACGCGGCTGCAGCAGGAGCTGGAC  991

Qy     102  AspLeuAlaAlaAspAlaValMetGlnAlaValMetGluMetGlnAlaAlaLysGluLeuGlnProLysSer  121
                ||||       |||||:::     ||||||||||||        |||||||:||||  :::
Db     992  GACCTGCTGCTGCAGGACCTGGACCTGGACCGGCCCCCAGAGCGCCGGCTGCAACCTGGAGAAGAAGCAG  1051

Qy     122  GlyAsnAlaPheIleAspMetLeuAsnGlyAsnGlyIlePheIleProIleGlySerIleArg  141
                ::::||                        :::||
Db    1052  --AAGAAGTTTGACCAGTCCTG------------------------------GCGGAGAGAAG  1084

Qy     142  GlyLeuGluAspAlaIleArgThrArgAspMetGluAsnThrAspProSerGluGln  161
                :::                 ::::||||
Db    1085  ACCATCTCTGCCAAGTATGCAGAGAGCCGCCGACCGGGCTGAGGGGAGCCCCAGAGAAG  1144

Qy     162  IleAlaLysAlaValMetAspLysPheGlnIleIleLeu--ProGlyLeuValAlaAlaSer  181
                |||||||                            ||  ||||||
Db    1145  GAGACCAAGGCCTCTGTCG--------------------------CTGGCCGGG---  1172

Qy     181  nMetIleAlaGlyLysAsnProPheLysMetProGlnMetProGlnLysAlaGlnAlaAla1  201
                ::::                          ||:|||||||
Db    1173  -------------CCCTGGAGGAAGCCATGACGCAGAGAGGCGGACTGGAGCCGGC  1213

Qy     201  aProSerValPheGlnAlaLeuGln----------------------------------Ar  213
                |                ::::
```

```
Qy     430  ------------PheAlaProArgLysIleProThrArgProArgLysMetLeuProL  445
                           |||                            :::     ||||||
Db    5761  GGATGACGAGCGGAGGAACGCCGAGCAGTACAACGGACCCGCACACAAGGCATCTACCCG  5820

Qy     445  euLeuIleGlySerAspProLysValGlnGluGlyLeuIleArgArgHisProSer------  462
                  :::                           :::|||       ||||||
Db    5821  CCTGAAGCAGCTCAAGCGGCAGCTGGAGGAGGCCG--AAGAGGAGGCCCAGCGGGCCAA  5877

Qy     462  --------------------------------------------------------------  462
Db    5878  CGGCTCCGCCGCGGAAACTGCAGCGGCCGAGCTGGAGGACGCGCCACTGAGACGGCCGATGCCAT  5937

Qy     462  --------------------------------------------------------------  462
Db    5938  GAACCGGCGAAGTCAGCTCCCTAAAGAACAGAGCTCAGGCCGCGGGAGACCTGCCGTTTGTGT  5997

Qy     463  ----ThrGluTrpLys------------------------IleAlaL  469
                :::                                     ::|||
Db    5998  GCCCCGCCGAATGCCCGGAAAGGCGCCGGGATGGCTCCGACGAAGAGGTAGGATGGCCAA  6057

Qy     469  ysGluSerArgValLeuThrAsnLeuLysAsnProSerLeuAlaAlaLeuPheMetA  489
                |||:::        ||||||:||
Db    6058  AGCGGATGGGCGTGGAGGCCGGAACCTGCCAGAATAAGCCTTCTTCCTGCAGCCTGAGATGG  6117

Qy     489  spAspLysLeuGluAsnThrLeuLysGlyArgLeuGlnMetLeuThrAspGluGlnLysGlyGlyA  509
                ::: :::  ||||||                        |||
Db    6118  ATGACAGAGCAACACCACCAGCCTCCCCTTCCCAG-----ACCCGCAG-------------  6161

Qy     509  rgThrArgValLysThrIleArgLeuProArgLeuPheGlyAla----------------  524
                :::|||||                            :::
Db    6162  --------CACGCCTCTCCCCACCTTCTTGGGACTGCTGTGAACATGC  6201

Qy     525  --ProThrAlaLysGlyAlaMetIleAlaAspAlaLysValPheGlnAspIleGluGluArgP  544
                   ||||||                            |||
Db    6202  CTCCTCCTGCCCTC---------------------------------CGCC  6219

Qy     544  roIleProProLeuPheGluProLysGlyProGlyArgHisThrArgLeuArgTrpThrGlyA  564
                ||:::|||||||                        |||
Db    6220  CCGTTCCCCCATCCGTTCCCT-----------------------------  6242

Qy     564  laAsnGluLysGlyIleLeuProGlyLeuGlySerArgPheIleLeuProSerLeuAspProT  584
                |||:::||||||                        :::||
Db    6243  --------------CCAGGTGTTGTTGAGGGCATTTGGCTTCCTCTGCTGCATCCC-  6284

Qy     584  hrMetProAlaLeuAsnThrAlaPheSerThrGlnGlyArgAlaArgAspGluTrpAspT  604
                ::::||||||       :::
Db    6285  --CTTCCAGCTCCCTCCCCTGCTCAGAATCTG-----------------  6314

Qy     604  hrMetPheLysIleProAsnAsnTrpAsnProGly  615
                |||||||                    :::
Db    6315  ---------ATACCAAAGACAGAGCAGGGCCCGGGT  6338
```

RESULT 11
US-10-264-049-543

; Sequence 543, Application US/10264049
; Publication No. US2004005579A1
; GENERAL INFORMATION:
;   APPLICANT: Birse et al.
;   TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;   FILE REFERENCE: PA133P1
;   CURRENT APPLICATION NUMBER: US/10/264,049
;   CURRENT FILING DATE: 2002-10-04
;   PRIOR APPLICATION NUMBER: PCT/US01/18569
;   PRIOR FILING DATE: 2001-06-07
;   PRIOR APPLICATION NUMBER: US 60/209,467
;   PRIOR FILING DATE: 2000-06-07
;   NUMBER OF SEQ ID NOS: 4360
;   SOFTWARE: PatentIn Ver. 3.1
; SEQ ID NO 543
;   LENGTH: 4041
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:

Db 1214 T-----------CAACAAGCAGTTCCGCACGGAGATGAGGACCTTATGAGCTCCA 1258

Qy   213 gAlaMetLeuGlyLysAsnAlaProValAlaGly------------GlyAr   226
          |||::: ||||||
Db 1259 AGGATGATGTGGGCAAGAGTCTCCACGAGCTGGAGAAGTCCAAGCGGGCCCTAGAGCAGC 1318

Qy   226 gGly-GluGlnArgMetMetAsnArgValAspArgMetGln-----------   242
          ||| ||||| ::: :::::::::
Db 1319 AGGTTGGAGGAGATGAAGACGCAGCTGGAGAAGACGCAGCTGCAGGCCACCGAAG 1378

Qy   243 -----------------------------GlnArgGluL   246
                                       :::|||:::|||
Db 1379 ATGCCAAGCTGCGGGTTGGAGGTCAACCTGCGAGGCCATGAACAGGCCCAGTTCGAGCGGGACC 1438

Qy   246 euGlnGluGluAspGluAspAspLeuGluGluAspGluAspAspValProArgArgS   266
          ::: ::: ::: ||| ::::::
Db 1439 TGCAGGGCCGGGACGCAGCAGCAGAGAAGAAGCAGCTGGTCAGACAGGTGCGGG 1498

Qy   266 erSerArgGlyGlnProGlnSerGluAlaGluHisGlnArgArgAspLeuAlaArgArgL   286
          ::: |||||| :: :::
Db 1499 AGATGAGGCAGAGAGCTGGAGGACGAGGAGCAGCGCTGCATGGCAGTGGCCCGCCCGGA 1558

Qy   286 euLysSerSerProArgLeuLysGluLeu------------------------   295
          :::::|||||:::|||
Db 1559 AGAAGTGGAGATGGACGACCTGGAGGCCGCACATCGACTCGGCCCAACAAGAACC 1618

Qy   296 -----------LeuGlnAsnAlaGluValGlnSerLeuLeuSerThrGlnA   309
                     ::: ::: :::
Db 1619 GGGACGAAGCCATCAAACAGCTGCGGGAAGCTGCAGGCCCAGATGAAGGACTGCGCGG 1678

Qy   309 rgMetArgAspSerProLeuSerArgArgAlaMetAsnAspGluAspGluS   329
          |||||| :::::::
Db 1679 AGCTGGATGACACCCGCGCCTTCGTGAGGAGATCCTGGCCCAGGCCAAAGAGAACGAGA 1738

Qy   329 erAlaPheArgAlaMetGluAlaArgAlaLysLysAspLeuAsnLeuValLeuG   349
          ::::::: ::: |||||||
Db 1739 AGAAGCTGAAGACGAAGACCATGGAGGCC-----------GAGATGATCCAGTTGCAGGA-G   1782

Qy   349 lyLeuHisGlyPheGlySerAspAspAspGluGluAspGluGlnAsnLeuIleA   369
          ||| :::::: ::: |||||||
Db 1783 GAACTGCAGCTGCGCCAAGCGCCAGCCGCCAGCAGCGGGGATGGAGCTGGCTG 1842

Qy   369 spProMetAsnSerPheArgArgAlaProLeuArgLeuSer-----------   383
          ||| :::
Db 1843 ACGAGATCGCCAACAGCAGCCGCAAAGGAGCCCTGCGGTTAGAGGAGAAGGCGGCGTTGG 1902

Qy   384 --------------------SerGlyPheGluLysLeuLysSer---------   392
                              |||||:::
Db 1903 AGGGCCGCATCGCCAGCGTCGGAGGAGGAGCTGGAGGAGGCAACACGGAGCTGA 1962

Qy   393 --AsnAspGluLeuLysSerAla-------------------------   399
            :::||| :::
Db 1963 TCAACGACCAGCGCTGAAGAAGGCCAACCTGCGAGATCGACCAGATCAACAACCGACCTGAACC 2022

Qy   400 -----------------------LeuAspArg-------------------   402
                                 |||:::|||
Db 2023 TGGAGGCGCAGCCACGCCCAGGAGAAGACGAGAATGCTCGGCACGCTGGAACGCCAGAACA 2082

Qy   403 -----IleLysTyrArgValAspGlu-----------LysTyrLeuAla-   415
               :::||| :::::::
Db 2083 AGGAGCTTAAGGTCAAGCTCGAGGAGATGGAGGGCACTGTCAAGTCCAAGTACAACAGGCCT 2142

Qy   416 ProLysPro-------------------MetGluPheAsnProLysPro   425
          |||:::|||
Db 2143 CCATCACCGCCCTCGAGGCCCAAGATTGCACAGCTGGAGGAGCAGCTGGACAACGAGACCCA 2202

Qy   426 --------GlnProGlyTyr-----------------   429
                  :::|||
Db 2203 AGGAGCGCCAGGCAGCCTGCAAACAGTGCGTCGGACGAGAAGAGAGTGAAGGATGTGC 2262

Qy   430 -----------PheAlaProArgLysIleProThrArgProArgLys   441
                     ||||||:::
Db 2263 TGCTGCAGGTGATGACGAGCGGAGGAACGCCGGAGCAGTACAAGGACCAAGGGCCGACAAGG 2322

Qy   442 MetLeuGlyLeuIleGlySerAspProLysProValGlnGluIleArgArgHisPro 461
          |||:::   ::: :::   ---
Db 2323 CATCTACCCGCCTGAAGCAGTCCAAGCGGCCAGCTGGAGGAGGAGGCCG---AAGAGGAGCCC 2379

Qy   462 Ser---------------------------------------   462

Db 2380 AGCGGGCCAACGCCTCCCGCCGGAAACTGCAGCGCGGAGCTGGAGGACGCCACTGAGACGG 2439

Qy   462 --   462
Db 2440 CCGATGCCATGAACCGCGAAGTCAGCTCCCTAAAGACAAGCTCAGGCGCGGGGGACCTGC 2499

Qy   463 ------------------------ThrGluTrpLys------   466
                                  :::
Db 2500 CGTTTGTCGTGCCCCGCCGGAATGGCCCGGGAAAGGCGCCGGGGATGGCTTCCGACGAAGAGG 2559

Qy   467 ---IleAlaLysGluSerArgValLeuThrAsnLeuLysAsnProSerLeuAlaAla 485
             :::   ::: :::
Db 2560 TAGATGGCAAAGCGGATGGGCGTGAGGCCAAACCTGCCGAATAAGCCTCTTCCTCCGCAG 2619

Qy   486 LeuPheMetAspAspLysLeuGluAsnThrLeuLysLysGlyLysArgGlnMetLeuThrAspGlu 505
                  ::: |||
Db 2620 CCTGAGATGGATGGACAGACAGACACCACCAGCCTCCCCTTCCCAG----ACCCCGCAG 2673

Qy   506 GlnLysGlyArgThrArgValLysThrIleArgAlaLeuProArgLeuPheGlyAla--- 524
                              ||| :::::::
Db 2674 -------------------------CACGCCCTCCTCCCCACCTTCTTGGGACTGCT 2703

Qy   525 --------------------ProThrAlaLysGluMetIleAspAspLysValPheGlnGlnAspIle 540
                              :::::: |||
Db 2704 GTGAAACATGCCTCCTCCTGCCCTC----------------------------- 2727

Qy   541 GluGluArgProIleProGluPheGluProLysGlyArgHisThrArgLeuArg 560
          :::::::
Db 2728 ---------------CGCCCCGACCCCCATCCCGTTCCCT----------- 2754

Qy   561 TrpThrGlyAlaAsnGluLysGlyLeuIleProGlyLeuGlySerArgPheIleLeuProSer 580
                              ::: :::
Db 2755 ----------------------CCAGGTGTTGTTGAGGGCATTGGCTTCCTCTG 2787

Qy   581 LeuAspPro---------------ThrMetProAlaLeuAsnThrAlaPheSerThrGlnGly-Ar 597
          ::: |||
Db 2788 CTGCATCCCCTTCCAGCTCCCTGCTCAGAATCTGATACCAAAGACAGGGCCCG 2847

Qy   597 gAlaArgAspGluTrpAspThrMetPheLysIleProAsn----------- 610
          |||:::||| :::
Db 2848 GGCCCAGCGAGAGCGACCAGGCTTCCTCCTTGCCAAAAGCACACAGAT 2907

Qy   611 ---------------------AsnTrpAsnProGluValGlyMetLysMetAsn 623
                               :::
Db 2908 GTTGAGGCGGAGCAGGGCAGGCCCCGGGAGGGGCCAGAGTTTCTATGAAT 2959

```
;   TYPE: DNA
;   ORGANISM: Caenorhabditis elegans
US-10-369-493-30630

Alignment Scores:
Pred. No.:           2.1e-05         Length:         3257
Score:               162.50          Matches:        132
Percent Similarity:  37.14%          Conservative:   99
Best Local Similarity: 21.22%        Mismatches:     242
Query Match:                         Indels:         151
DB:                                  Gaps:           28
                     4.94%
                     17

US-10-736-868-2 (1-643) x US-10-369-493-30630 (1-3257)
```

```
Qy     18 LeuSerAlaGlnSerProSerThrSerAspAlaProGlyAlaLeuLeuSerSerVal 37
             :::  ||||||                       :::        ||||||
Db    859 ATCCGAAGCTCAATATGAGAGACTTCAAGAGACTTAGCTACTCTTAAGACACTGTGTG 918

Qy     38 GlyLysSerHisGlnLysGluProLeuAlaProSerMetGluAlaLeuGluLeuMetGly 57
             :::   ||||||    ||||||||||:::        |||   :::
Db    919 CAAGAGGAGGAGAAGAGCGCCACTC------CAAGAGGGTGCGAACGTCTCAAC 969

Qy     58 ValGlnPheValAspAlaLeuileLeuLysGlyGlnMetAlaLysGlyAlaPhe 77
             ||||||    :::|||  |||||:::|||       :::
Db    970 AAAGAGACCGGCTGACCCTTG--------GCGCAATTGGAGGCTAGCCAAGGGAAGCACC 1020

Qy     78 LysThrGlnLeuGluValLeuGluLysValHisProAspGlnPhe----------- 92
             |||    :::      :::             :::
Db   1021 CGGCGAAGTTGAGGAAAGAACCCATGAATGAGACAGAAGGTTGCCCTGAAGGAAAG 1080

Qy     93 -----AspLysTyrLysLeuLysValAspAspAlaLeuAlaAspAlaValMetGln 110
               |||       |||||||||:::|||||||||::::::        ||| |||
Db   1081 CTTGCCGATGCCAAGAAGTTGAAGTTGAAGAAGTCGTGCCGTTGAAATCAACAAG 1140

Qy    111 Gln---------AlaGluMetAlaLysLeuGlnLysProLysSerGlyAlaAsnAlaPhe 125
             |||                :::|||   ||||||         :::
Db   1141 CAAAAGAAGCTTGTCGAGGCCGAATGCCTGACTTGAAGAAGAACTGCCAAGATGTTGAC 1200

Qy    126 IleAspMet------LeuAsnGlyIleProlleGlySerSerileArgGlyLeu 143
             |||  :::               ||||||        ||||||||||
Db   1201 CTCTCATTGGAGAAAGGTTGAGGCGAAGAAGAAATGCCAAGGAGCACCAGATTCGTGCTCTT 1260

Qy    144 GluAspAlaIleArgThrGlnArgAsp---------MetGluAsnThrAsp 157
             :::|||  ||||||       |||          |||   :::
Db   1261 CAAGATGAAATGAACACACGAGCAGACACATCTCTAGCTTAACAGGAGGCCAAGAAC 1320

Qy    158 ProSerGluGlnIleAlaLysPheGlnThrGlnIlelleLeuProGly 177
             |||||||    :::|||    |||||||:::|||:::
Db   1321 CAAGAGAAGCAGAACAAGAAACTCACCCGAGGAT--CTCCAAGCTGCTGAGGAACAAAAC 1377

Qy    178 LeuValAlaAsnMetIleAlaGlyLys----------- 186
             |||||||    :::
Db   1378 TTGGCAGCCAACAAGCTCAAGGTCAAACTCATGCAATCTCTTGAAGACTCTCGAGACACC 1437

Qy    187 ----------AsnProPheLysMetProGlnMetArgLysArgLysAlaGlnAla 200
                        |||   ||||||:::           ||||||||||:::
Db   1438 ATGGAGGCCGAGAGAGACAAACCGTGCTGATATGGACAAGAATAAGAGAAGGCCGGA--- 1494

Qy    201 AlaProSerValPheGlnAlaLeu-----------Ala 211
             ||||||||||    |||  |||               |||
Db   1495 GGAGAGCTCAAGATTGCTCAAGAAACCTTGGAAGAGCTCAACAAGTCAACAAGAGCGATCC 1554

Qy    212 GlnArgAlaMetLeuGlyLysAsnAlaProVal----- 229
             ||||||||||      |||||||||:::
Db   1555 GAAAATGTCTTTGACCGCCAACCGAATTGCCACACCTTGGGAATGAAACTCGAAGAC 1614

Qy    230 GlnArgMetMetAsnArgValAspGlnMetGlnArgGluGlnGluGlu 249
             ||||||||||:::|||||| :::  :::  ||| :::
Db   1615 GAGCAAGCTCTGTTGCTAAACTCCAAAAGGGAATTCAGCAAGGCGTAGGAGTCAAG 1674

Qy    250 AspGluAspAspAspLeuGluAspValProArgValProArgArgSerAspGly 269
             ||||||||||:::|||||||||:::
Db   1675 GATCTT---CATGATCAACTTGCCGATGAGAAGAGATGCTCGCCATGCGTCG 1731
```

```
Qy    270 GluProGlnSerGluAlaGluHisGlnArgArgAspLeuAlaArgLeuLysSerSer 289
             :::|||
Db   1732 AGAGCCGACCAACAGGCTGAGTACCAT-----GAGCTCACCGAGCAACTCGAAGACCAA 1785

Qy    290 ProArgLeuLysGlyLeuLeuGlnAsnAlaGluValGlnSerLeuLeuSerTyrGlnArg 309
             |||                                 :::        :::
Db   1786 GCCCGT------------------GCCACCGCTGCCCAAATTGAGCTCGGAAAA 1821

Qy    310 MetArgAspProLeuSerLysArgArgProLeuAlaMetAsnAspGluAspGluSer 329
             :::|||  ||||||:::          |||  |||||||||
Db   1822 AAGAGGAGGATGCTGAGCTTACCAAGCTCCGT--CGTGATCTTGAAGAGTCT 1869

Qy    330 AlaPheArgAlaMetGluAlaArgAlaLeuAspAlaLysGlnLysSerGlnLeuValLeuGly 349
             |||||||          :::||||||
Db   1870 GGACTTAAATTCGGAGAGCAACTCCACTGTTCTCAAGAGAAG-----GGA 1914

Qy    350 LeuHisGlyPheGlyGluSerAspAspAspAspGluGluAspGluAsnLeulleAsp 369
             |||
Db   1915 TCAGATGCTATTCAAGAGAGCTTCGATCAAATGCAGCAA------------- 1953

Qy    370 ProSerGluAsnSerPheArgArgAlaProLeuArgLeuSerGlyPheGlyValGluLys 389
             ||| :::        :::|||||||
Db   1954 -------------------CTTCAAAGCCAAAAGGGAAGAATTGAGAAG 1983

Qy    390 LeuLysSerAsn----------AspGluLysSerAlaLeuAspArg---Ile 403
             |||||||                |||||||||
Db   1984 GAGAAAGGACATATCCAACCGCGAGTTCGATGAGTCCTCTGCTCTTGACCAAGAGCC 2043

Qy    404 LysTyrArgValAspAsp---------ValGluLysTyrLeuAlaProLysProMetGlu-Ph 421
             |||||||                   :::                    :::
Db   2044 AAGCTCCGGCTGACCCAGGAGAGAATTGCCAAGGGATACGAGGTTCAGACTTCTTGAGCTT 2103

Qy    421 e-----------AsnPro----------LysProGlnProGlyTy 429
             :::          |||
Db   2104 CGTCTTCAAGGCTGACGAACAATCCCGCGACTTCAAGATTTCGTCCTCAAAGGGACGC 2163

Qy    429 rPheAlaProArgLysIleProThr----------ArgProArgGly 441
                        |||                   |||
Db   2164 CTCAACTCCGAGA--ACTCCGATTTGGCCGCCAAGTCCAAGATC 2220

Qy    441 sMetLeuProLeuLeuIleGlySerAsp---ProLysValGlnGlulleArgArgHi 460
             ||||||                     ||||||
Db   2221 CAAGCTGCCAACCGCCTCAAACTCCAATTCTCCAACGA--TTGGATCATGCCAAGAGACA 2279

Qy    460 sProSerThrGluThrProLysTrpLysValLeuSerArgValLeuThrAsnLeuLysAsnAs 480
             :::         :::|||||||
Db   2280 GGGCCGGAGGAA----------TCCCGCGAGAGACAAAATCTTTCCAATTTGTCGAAGAA 2330

Qy    480 nProSerLeuAlaLeuPheMetAspArgLeuAsnThrLeuLysGluAsnThrLeuLysGly1 500
             |||||||:::
Db   2331 CTTTGGCTCCGTCGAGCTTGAACAGCTCAAGGAGACGAGAGTCCCTGGAAAGAA 2390

Qy    500 nMetLeuThrAspGlu----------GlnLysGlyArgThrArgVa 512
             |||||||||||
Db   2391 CGAGGCTTCTCGTCAACTTCCAAGGCTTCCGTTGAACTCGATCAATGAGAACTAAGTT 2450

Qy    512 lLysThrIleArgAlaLeuProArgLeuPheGlyAlaProThrAlaLysGluMetIl 532
             :::|||
Db   2451 CGAGACC----------------GAGGGACTCAT 2468

Qy    532 eAspAlaLysValPheGlnAspIleGluArgProIleProProLeuProPheGluPr 552
             :::|||
Db   2469 TGGAGCCGACGAGTTCGATGAGGTCAA-GAAACGCCAAAACCAGAGACTTCGGAGATCC 2527

Qy    552 oLys 553
             :::|||
Db   2528 AAGA 2531
```

```
RESULT 13
US-10-956-157-859
; Sequence 859, Application US/10956157
; Publication No. US20050118625A1
```

; GENERAL INFORMATION:
; APPLICANT: Wyeth
; APPLICANT: Mounts, William
; TITLE OF INVENTION: NUCLEIC ACID ARRAYS FOR DETECTING GENE EXPRESSION ASSOCIATED WITH
; TITLE OF INVENTION: HUMAN OSTEOARTHRITIS AND HUMAN PROTEASES
; FILE REFERENCE: 031896-043000 (AM 101081)
; CURRENT APPLICATION NUMBER: US/10/956,157
; CURRENT FILING DATE: 2004-10-04
; NUMBER OF SEQ ID NOS: 319805
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 859
; LENGTH: 5883
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-956-157-859

Alignment Scores:
Pred. No.:   4.67e-05
Score: 162.50
Percent Similarity: 35.90%
Best Local Similarity: 21.79%
Query Match: 4.94%
DB:

Length:	5883
Matches:	136
Conservative:	88
Mismatches:	213
Indels:	192
Gaps:	25

US-10-736-868-2' (1-643) x US-10-956-157-859 (1-5883)

```
Qy    12 GlyPheCysIleAlaProLeuSer------AlaGlnSerProSerThrSerAsp------  27
            ||||||:|||    ::|||   ::|
Db  3860 GGCTTCTCAGCCAGTCGACACAGCAAGTCCAGCAAGCTCACCAAGGACTTCTCCGGCCTGG 3919

Qy    28 AlaProGly-------AlaLeuLeuSerLeuValGlyLysSerHisGlnLysLeuPro     45
            :::||||||            |||||||:||   ::  :||||||  ||||||:|||
Db  3920 AGTCCCAGTCTGCAGGACACTCAGGAGCTG--CTGCAGGAGGAGAACCGGCAGAAGTGAGC 3978

Qy    46 LeuAlaProSerMetGluAlaLeuGlu------------------------------     54
            |||:|||||        ::|||
Db  3979 CTGAGCACCCAAGCTCAAGCAGCAGGTGGAGGAGGACGAGAGAATTCCTTCCGGGAGGAG 4038

Qy    55 -----------------------------------LeuMetGlyValGlnPhe      60
                                                  ||:
Db  4039 GAGGAGGAGGAGGAAGCAGGGCCCAGCAACCTGGAGAGCAGCAGATCGCCACCTCCATGCCCCAGGTG 4098

Qy    61 ValAspAlaLeuIleLysLysGlyGlnMetGluMetAlaLysAlaPheLysThrGln     80
            :::|||||||||    ||||||:||   :::  :||||||||
Db  4099 AAGGTGGCCGCCGTACGACACAGACAGTGTGGGTGCCTGGAAACTGCT 4149

Qy    81 LeuGluValLeuGluLysAlaProLeuGlu------------------------------     90
            |||||
Db  4150 GACCCTGGTGCCTGAGGAGGAAGCTTCCAGAGACCGTGCACCAGGGCCTGAGCAGCAGGAG 4209

Qy    91 --GlnPheAspLysTyrLysLeuLys-------------------------ValAsp    101
            ||||||::|||  ||||||
Db  4210 AAGGTGGCCGCCTACGACACAGACAGTGGAGAAGACGCGGCCAGAGAGCTGCAGCAGGAGGAC 4269

Qy   102 AspLeuAlaAlaAspAlaMetGlnGlnAlaLeuAspMetAlaLysAlaAspProLysSer    121
            ||||||       ||:::|||:||||||:||||||
Db  4270 GACCCTGGTGCCTGAGGAGCCGTCGACCTGGAGGAGAAGAAGCAG 4329

Qy   122 GlyAsnAlaPheIleAspMetLeuAsnGlyAsnGlyIleProIleGly-------------    137
            |||||||
Db  4330 AAGAAG-------TTTGACCAGTCCTCTGGCGGAGGAGAAGACCATCTCTGCCAAGTATGCA 4383

Qy   138 -----------------------------------SerSer    139
                                                  |||
Db  4384 GAGGAGGCGGACCCGGCTGAGGCGGAGGCCCAGAAGGAGAAGCAAGGCTCTGTCGCTG 4443

Qy   140 IleArgArgGlyLeuGluAspAlaAlaIleArgThrGlnAsnThrAspProSer    159
            |||||||||||:::||||||||||  ::|||||||
Db  4444 GCCCCGGCCACCTGGAGGAGAAGGCCCATGAGGAGGCGGAGCGGCTCAAC------  4497

Qy   160 GluGlnIleAlaLysAlaValMetAspLysPheGlnThrGlnIleLeuProGlyLeuVal    179
            ::::|||||:|:||
Db  4498 -----------------------------AAGCAGTTCCGCACGCAG------------ATG 4518
```

Qy   180 AlaAsnMetIleAlaGlyLysAsnProPhe---LysMetProGlnMetArgLysAla 198
            ||:::| |||| :::||| |||
Db 4519 GAGGACCTTATGAGCTCCAAGATGATGTGGCCAAGAGTGTCCAGCTGGAGAAGTC- 4577

Qy   199 GlnAlaAlaProSerSerValPheGlnIleAlaLeuAlaLeuArgAlaMetLeuGlyLys 218
            ||||||:|| ||:|||:::|||||||
Db 4578 CAAGCCGGGCCCT------------------------------AGAGCAGCAGGTGGAGGA 4607

Qy   219 Asn--------AlaProValAlaGlyThrGlyGluGlnArgMetMetMetAsn 235
            ::: ||||||| :||:||||
Db 4608 GATGAAGACGCAGCTGGAAGAGCTGCAGGCCACCAAGGATGCCAAGCT 4667

Qy   236 -ArgValAspGlnArgMetGln------------GlnArgGluGlnLeuGluGl 249
            ::::||| ::||||
Db 4668 GCGGTTGGAGTCAACCTGGACCCATGAAGGCCCCAGTTCGAGCGGCGGACCTGCAGGGGCCG 4727

Qy   249 uAspGluAspAspAspAspLeuGluAspValProArgArgSerAspGl 269
            |:||:::   ::: :::   :::||:|||||
Db 4728 GGACGAGCAGAGCGAGGAGAAGAAGCAGCTGGTCAGACAGGTGCGGGAGGATGGAGGC 4787

Qy   269 yGluProGlnSerGluAlaGluHisGlnArgArgAspArgAspArgGluLeuLysSerSe 289
            ||| ::: :::||||||| :::
Db 4788 AGAGCTGGAGGACGAGAGGAGCAGCGCTCGATGGACGGTGGCCGCCCGGAAGAAGTGGA 4847

Qy   289 rProArgLeuGlyGluLeu 295
            ||||:|||
Db 4848 GATGGAAGGACCTGAAGGACCTGGAGGAGGCCGACCATCGACTCCGGCCACAAGACCGGCAGAAGC 4907

Qy   296 ----------LeuIleAsnAlaGluValGlnSerLeuLeuSerGlnLysTyrGlnArgMetArgAs 312
            ::: ||||:|||||||||
Db 4908 CATCAAACAGCTGCGGAAGCTGCAGCTGCGGGCCCAGATGAAGGAGCTGCATGCGGCGCTGAGGAC 4967

Qy   312 pSerProLeuSerLysArgArgProLeuAlaMetAsnAspGluSerAlaPheAr 332
            |:||| |||||||
Db 4968 CACCCGGCCTCGTGAGGAGAGATCCTGGCCTGGCCAAGAGCAAAGGAGAAGCTGAA 5027

Qy   332 gAlaMetGluAlaArgAlaAlaLysLeuAspAspLysGlnLeuSerGlnLeuValLeuGlyLysHisGl 352
            :::||||||
Db 5028 GAGCATGGAGGCC-------------GAGATGATCCAGTTGCAGGAGGAACTGGCCAGC 5072

Qy   352 yPheGlyGluSerArgAspAspAspAspProLeuGluGluGlyGluAspSerProSerGl 372
            ||| ::||||||||| :::
Db 5073 CGCGGAGCGTGCCAAGCCCAGGCGCCGGATGAGCTGGCTGCTGCGAGATCGC 5132

Qy   372 uAsnSerPheArgArgAlaProLeuArgLeuSer------------------ 383
            |||||| ::|||||||||
Db 5133 CAACAGCAGATTGAGGAGGCCTGGCGTTAGAGGAGGAAGCGGCGTCTGGAGGCCCCAT 5192

Qy   384 -SerGlyPheValGluLeuLysLysSer---------------------AsnAspGl 395
            ::: |||||
Db 5193 CGCCCAGTGCTGGAGGAGGAGCTGGAGGAGGAACGGCAGCAGCAACGAGCTGATCAACGACACG 5252

Qy   395 uLeuLysSerAlaLeuAspArgIleLysTyrLysArgValAspAspValGluGluTyrLeuAl 415
            ||||||| |||||||
Db 5253 GCTGAAGAAGGCC--------------------------AACCTGCAGATCGACCAGATC- 5286

Qy   415 aProLysProMetGluPheAsnProLysProGlyTyrPheAlaProArgLysIl 435
            ||| :::|||||| ||||||
Db 5287 ----------------------------AACAC-CGACCTGAACCTGA- 5306

Qy   435 eProThrArgProArgLysMetLeuProLeuLeuIleGlySerAspProLysValGlnIl 455
            :::|||||||| ::::|||
Db 5307 ----GCCGCCACGCGCC-AGAAGAACGAGAAATGCTCGGCAGCAGCCCAGACCTGGAACCCAGAC 5361

Qy   455 uGluIleArgArgHisProSerThrGluTrpLysIleAlaLysGluSerArgValLeuTh 475
            :::|||||| |||||||
Db 5362 AAGGAGCTTAAGGTCAAGCTGCAGGAGATGG----------AGGGCACTGTC 5403

Qy   475 rAsnLeuLysAsnAsnProSerLeuAlaLeu------PheMetAspAs 490
            |||||::: :::||||||||||
Db 5404 AAGTCC--AAGTACAAGGCCTCCATCACCGCCCCTCGAGCCCAAGATTCCACAGCTGAGGA 5462
```

```
Qy    490 pLysLeuGluAsnThrLeuLysArgGlnMetLeuThrAspGluGlnLysGlyArgTh 510
              :::||||||:||    |||    ||||||:|
Db   5463 GCAGCTGGACAACGACCAAGGAGACCAAGGAGCGGCCAGCCGCCTGC---AAACAGGTGCCTCGGAC 5519

Qy    510 rArgValLys 513
              ||| ||:
Db   5520 CGAGAAGAAG 5529
```

RESULT 14

US-10-283-975A-279

```
; Sequence 279, Application US/10283975A
; Publication No. US20040110792A1
; GENERAL INFORMATION:
;   APPLICANT: Ortho-Clinical Diagnostics, Inc.
;   TITLE OF INVENTION: Methods For Assessing and Treating Leukemia
;   FILE REFERENCE: CDS 293 PCT
;   CURRENT APPLICATION NUMBER: US/10/283,975A
;   CURRENT FILING DATE: 2002-10-30
;   PRIOR APPLICATION NUMBER: 60/340,938
;   PRIOR FILING DATE: 2001-10-30
;   PRIOR APPLICATION NUMBER: 60/338,997
;   PRIOR FILING DATE: 2001-10-30
;   PRIOR APPLICATION NUMBER: 60/340,081
;   PRIOR FILING DATE: 2001-10-30
;   PRIOR APPLICATION NUMBER: 60/341,012
;   PRIOR FILING DATE: 2001-10-30
;   NUMBER OF SEQ ID NOS: 900
;   SOFTWARE: PatentIn version 3.1
; SEQ ID NO 279
;   LENGTH: 3320
;   TYPE: DNA
;   ORGANISM: HUMAN
;   FEATURE:
;     NAME/KEY: misc_feature
;     LOCATION: (1)..(3320)
;     OTHER INFORMATION: N=any base
US-10-283-975A-279
```

Alignment Scores:

```
Pred. No.:                          Length:        3320
Score:                  2.42e-05    Matches:        160
Percent Similarity:       38.26%    Conservative:   135
Best Local Similarity:    20.75%    Mismatches:     251
Query Match:               4.92%    Indels:         228
DB::                          19    Gaps:            39
```

US-10-736-868-2 (1-643) x US-10-283-975A-279 (1-3320)

```
Qy     15 IleAlaProLeuSerAlaGlnSerProSerThrSerAspAlaProGlyAlaLeuLeuSer 34
              :::||||||||:::||||      |||   :::
Db    656 GTGGCGTCCCTCAGTTCCCAGCTCCCAGGACACCCAGGAGCTG----------------- 697

Qy     35 SerLeuValGlyLysHisGlnLysLeuProLeuAlaProSerMetGluAlaLeuGlu 54
              ||||||||:   ||||||||||      :::||:
Db    698 ---CTTCAAGAAGAAACCCGGACAGAAGCTCAACGTGCTCACGAAGAGCTGCGAGCTGGAG 754

Qy     54 Ser 54
              
Db    755 GAGGAGCCGGAACAGCCTGCGAACCAGCCTGGACGAGGAGATGGAGGCCAAGCAGAACCTG 814

Qy     55 -------------LeuMetGlyValGlnPheValAspAlaLeuIleLysGlyGlyGln 69
                           :::||||   ||||:::   |||||||
Db    815 GAGCCGCCACATCTCCACTCTCAACATCCAGCTCTCCGACTCGAAG--AAGAAGCTGCAG 871

Qy     70 -------------MetGluMetAlaLysGlyAlaPheLysThrGlnLeu 81
                           ::||:||   ||||||:::||||||
Db    872 GACTTTGCCAGCACCGTGGAAGCTCTGGAAGGGGAAGAGGAGGTTCCAGAAGAGAGATC 931

Qy     82 GluValLeuGluLysValHisProAspGln-------PheAspLysTyrLysLysLeu 98
              ||||   |||   :::    |||    :::||||
Db    932 GAGAACCTCACCCAGCAGTACGAGGAGAAGGCCGGCCGCCGATCGCCCGGATCGCCGAGGAG 991

Qy     99 Lys----------ValAspAspLeuAlaAlaAspAlaValMetGlnGlnAla 112
              |||           :::||||
```

```
Db   1958 GAGCTGAGGAGGAGCAGGG-CAACATGAGGAGGCCATGAGCGAGGCCGGTCC-GCAAGCCA 2015
Qy    433 ArgProArgLys------                    ArgProArgLys 441
          :::|:::
Db   2016 CACAGCAGGCCAGCCGCAGCTCAGCAACGAGCTGGCCACAGAGCGCAGCACCGGCCCAGAAGA 2075
Qy    442 Met---LeuProLeuLeuIleGlySerAspProLysValGlnGlyGluIleArgArgHis 460
          |||:||
Db   2076 ATGAGAGTTCCC--                   GGCAGCAGCTTGAGCGGCAGAACAAGAGCTCCGAGCA 2126
Qy    461 ProSerThrGluTrpLysIleAlaLysGluSerArgValLeuThrAsnLeuLysAsnAsn 480
                                      :::|||
Db   2127 AGCTCCACGAGATGG--                   AGGGGGCCGTCAAGTCCAAGTTCAAAT 2168
Qy    481 PheMetAspAspLysLeuGluAsnThr 495
          ||||
Db   2169 CCACC-ATCGCGGCGCTGGAGCCCAAGATTGCACAGCTAGAGGAGCAGGTCGAGCACGAG 2227
Qy    496 LeuLysGlyArgGlnMetLeuThr 503
          ::|  ::::|||
Db   2228 GCCAGAGAAACAGGCGCCGCCACCAAGTCGCTGAAGCAGAAGAACAAGAGCTGAAGGAA 2287
Qy    504 AspGlu 506
                      |||||
Db   2288 ATCTTGCTCAGGTGAGCCAAGATGGCCAAGAGCAGTACAAGGAGCAGGCAGAG 2347
Qy    507 LysGlyArgThrArgValLysThrIleArgLeuProArgLeuProArgGlyAlaProThr 526
          |||||
Db   2348 AAAGGCAATGCCAGGGTCAAGCAGCTCAAG--       AGGCAGCTGGAGGAGGCAGAG 2398
Qy    527 AlaLysAlaGluMetIleAspAla---          LysValPheGlnAspIleGluGluArg 543
          :::::::::        :::::::::::::|||
Db   2399 GAGGAGTCCCAGCGCATCAACGCCACCGCAGGAGCTCAGCGGGGAGCTGGATGAGGGCC 2458
Qy    544 ProIleProLeuPheGluPheGluProLysGlyArgHis--       --ThrArg 558
          ||        ::::::
Db   2459 ACGGAGCGCAAC--                   GGAGGCCATGGGCTGGAGGTGAACGCACTCAAGAGCAAG 2509
Qy    559 LeuArgTrpThrGlyAlaAspAsnGlyLysGlyIleProGly-----   LeuGlySerArgPhe 576
          ::|||
Db   2510 CTCAGG--                   CGAGGAAACGAGAGACCTCTTTCGTTCCTTCTAGAAGGTCTGGAGGACGTAGA 2566
Qy    577 IleLeuProSerLeuAspProThrMetProAlaLeuAsnThr 590
                                    :::|||
Db   2567 GTTATTGAAAATGCAGATGGTTCTGAGGAGGAAACCGAGCACTGGAGAAGCAGACTTCAAT 2626
Qy    591 AlaPheSerThrGlnGlyArgAlaArg 599
          :::|||  :::
Db   2627 GGAACCAAGGCCAGTGAATAAGCAACTTTCTACAGTTTTGCACCACGGCAGAAAACCAA 2686
Qy    600 AspGluTrpAspThrPheLysIleProAsnTrpAsnProGlyAspGluValGly 619
          ||||
Db   2687 AAACCAAAAAACAAAACAAAAAAAAACCCAACAAC--   -AACCCCA- 2728
Qy    620 PheLysMetAsnSerLysThrLysArgPheVal 630
          ||
Db   2729 -----  --GAACAAAGCAAAACCCCAGCAGACTGTA 2755
```

Alignment Scores:
Pred. No.:
Score: 4.94e-05
Percent Similarity: 159.00
Best Local Similarity: 38.30%
Query Match: 20.52%
DB: 4.83%
                            9

SEQ ID NO 1602
  LENGTH: 3388
  TYPE: DNA
  ORGANISM: Homo sapiens
US-09-954-456-1602

Length: 3388
Matches: 150
Conservative: 130
Mismatches: 230
Indels: 225
Gaps: 35

US-10-736-868-2 (1-643) x US-09-954-456-1602 (1-3388)

```
Qy     15 IleAlaProLeuSerAlaGlnSerProSerThrSerAspAlaProGlyAlaProGlyAlaLeuLeuSer 34
          :::|||||||||::|||
Db   1270 GTGGCGGTCCCTCAGTTCCAGCTCCTTCCAGGACACCCAGGAGTTG-----   -- 1311
Qy     35 SerLeuValGlyLysSerHisGlnLysGlnLeuProLeuAlaProSerMetGluAlaLeuGlu 54
          ::: ||::::::
Db   1312 ---CTTCAAGAAGAAACCCGGCAGAAGCTCAACGTGTCTCACGAAGCTGCCGCCAGCTCGGAG 1368
Qy     54                                                              54
Db   1369 GAGGACGCGGAACAGCCTGCAGACAGCCTGCAGGAGGATGGAGGCCAAGGAAGAACCTG 1428
Qy     55                     LeuMetGlyValGlnPheValAspAlaLeuIleLysLysGlyGln 69
          :::|||||
Db   1429 GAGCGCCACATCTCCACTCTCAACATCAGCTTCCGGACTCTCCGAAG--   -AAGAAGCTGCAG 1485
Qy     70                     MetGluMetAlaLysGlyGlyAlaPheLysThrGlnLeu 81
          ||||||||
Db   1486 GACTTTGCCAGCACCGTGGAAGCTCTGGAAGAGGGGAAGAAGAGGTTCCAGAAGGAGATC 1545
Qy     82 GluValLeuGluLysValHisProAspGln---       PheAspLysTyrLysLysLeu 98
          |||  |||     :::
Db   1546 GAGAACTTCACCCAGAGCACAGTACCAGGAGAAGGGCCCGCTTATGATAAACTGGAAAAGACC 1605
Qy     99 Lys---                    ValAspAspSerLeuAlaAspAlaValMetGlnGlnAla 112
          ||       ||||||||||
Db   1606 AAGAAACAGGCTTCAGCAGCTGGAGCAGCCTGGTTGTTGATTTGGACAACCAGCGGCAA 1665
Qy    113 GluMetAlaLysLeuGlnLeuProLysSerGlyAsnAlaPheIleLeuAspAspMetLeuAsnGlyAsn 132
          ::::: |||:::
Db   1666 CTCGTGTCCAACCTGGAAAGAAGCAGAGGAAA----     -TTTGATCAGTTGTTAGCCGAG 1719
Qy    133 GlyIleProIleGlySer-----                                   139
          |||  |||
Db   1720 GAGAAAAACATCTCTCCAAATACGCCGATGAGGAGCACGAGCGAGGAAGCCAGG 1779
Qy    140                    IleArgGlyLeuGluAspAlaIleArgThrGln 150
          :::|||:::
Db   1780 GAGAAGGAAACCAAGGCCCTGCTTCCCTGGCTCGCGGGCCCTTTGAGAGGGCCTTGGAAGCCAAA 1839
```

RESULT 15
US-09-954-456-1602
; Sequence 1602, Application US/09954456
; Patent No. US20020115057A1
; GENERAL INFORMATION:
; APPLICANT: Young, Paul
; TITLE OF INVENTION: Process for Identifying Anti-Cancer Therapeutic Agents Using Canc
; TITLE OF INVENTION: Sets
; FILE REFERENCE: 689290-76
; CURRENT APPLICATION NUMBER: US/09/954,456
; CURRENT FILING DATE: 2001-09-18
; PRIOR APPLICATION NUMBER: US/60/233,617
; PRIOR FILING DATE: 2000-09-18
; PRIOR APPLICATION NUMBER: US/60/234,052
; PRIOR FILING DATE: 2000-09-20

; PRIOR APPLICATION NUMBER: US/60/234,923
; PRIOR FILING DATE: 2000-09-25
; PRIOR APPLICATION NUMBER: US/60/235,134
; PRIOR FILING DATE: 2000-09-25
; PRIOR APPLICATION NUMBER: US/60/235,637
; PRIOR FILING DATE: 2000-09-26
; PRIOR APPLICATION NUMBER: US/60/235,638
; PRIOR FILING DATE: 2000-09-26
; PRIOR APPLICATION NUMBER: US/60/235,711
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: US/60/235,720
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: US/60/235,840
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: US/60/235,863
; PRIOR FILING DATE: 2000-09-27
; NUMBER OF SEQ ID NOS: 2276
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 1602

```
Qy    478 LysAsnAsnProSerLeuAlaAlaLeu-----------PheMetAspAspLysArgPhe 492
              |||                    :::::|||||||||
Db   2773 AAGTTCAAGTCCACCATGCCGGCGGCTGGAGGCCCAGAGATTGCCACAGTCGGAGGCAGGTC 2832

Qy    493 GluAsnThrLeuLysGlyArgGlnMetLeuThr----------------------------- 503
              ||||                  :::|||
Db   2833 GAGCAGGAGGCCAGAGAGAAACAGGCAGCCAGCCACCAGTCGCTGAAGCAGAGAGACAAGAAG 2892

Qy    504 ------------------------------AspGlu----------------------------- 505
                                           ||||||
Db   2893 CTGGAGGAAATCTTGCTGCAGGTGAGGCCCAGATGGCCAAGATGGCCAGCAGTACAAGGAG 2952

Qy    506 ------GlnLysGlyArgThrArgValLysThrIleArgLeuProArgLeuPheGly 523
                 |||||||              |||||||||  ::::::
Db   2953 CAGGCGAGAGAAGGCAATGCCAGGTCAAGCAGCTCAAG-----------AGGCAGCTGGAG 3003

Qy    524 AlaProThrAlaLysAlaGluMetIleAspAla-----------LysValPheGlnIle 540
              :::::||||||||          |||:::||||||
Db   3004 GAGGCAGGAGGAGGTCCCAGCGCATCAACCGCCACCCGCAGGAAGCTGCAGCCGGGAGCTG 3063

Qy    541 GluGluArgProIleProPheGluProLeuPhePheGluProLysGlyArgHisThrArgLeuArg 560
              :::|||
Db   3064 GATGAGGCC-----------------------------------------------------AAG 3072

Qy    561 TrpThrGlyAsnGlu----------LysGluIleProGlyLeuGlyArgPheIle 577
              :::||||||           :::|||||||
Db   3073 ---ACGGGCAACGAGGCCATGGCCGTGAGTGAACGCACTCAAGAGCAAGCTCAGA 3129

Qy    578 LeuProSerLeuAspProThrMetProAlaLeuAsnThrAlaPheSerThrGlnGlyArg 597
                                  ::::||||||
Db   3130 GGGCCCCCC--------------CCACAGGA-AACTTCCAGTGATGCACCAGGCGAGG 3173

Qy    598 AlaArgAspGluTrpAspThrMetPheLysIle 608
              :::|||
Db   3174 AAACGAGACCCTCTTTCGTTCCTTCTAGAGGTC 3206

Search completed: July 18, 2005, 17:30:51
Job time : 2624 secs
```

```
Qy    151 ArgAspMetGluAsnThrAspProSerGluGlnIleAlaAlaValMetAspLysArgPhe 170
              ::::||||              :::::||||||||||
Db   1840 GAGGAAACTCGAGCGGGACCAAC---------AAAATGCTCAAAGCCGAAATGGAA------ 1884

Qy    171 GlnThrGlnIleLeuProGlyLeuValAlaAlaMetIleAlaGlyLysAsnProPhe--- 189
                                     :::::::||||||||       |||:::
Db   1885 ------------------------GACCTGGTCAGCTCCAAGGATGACGTGGGC 1914

Qy    190 LysMetProGlnGlnMetArgLysAlaGlnAlaAlaProSerSerValPheGlnGlnAla 209
              :::|||||||          :::||||||   :::::
Db   1915 AAGAACGTCCATGAGCTGGAGAAGTCCAAGCCGGCCCTGGAGAACCCAGATGAGGAGATG 1974

Qy    210 LeuAlaGlnAlaArgAlaMetLeuGlyLysLysAsnAlaProValAlaGlyLysGlyGluGlu 229
              |||               ||||||       ::||||||||||||||||
Db   1975 AAGACGCAGCTGGAAGAGCTGGAGGACGCAGCTGCCAAGCCTCG-----------GAGGAC 2022

Qy    230 GlnArgMetMetAsnArgValAspGlnArgMetGln------------------------Gln 243
              :::::::           |||||||
Db   2023 GCCAAACTG--------CGGCTGGAAGTCAACATGGCCAGGCGGCTCCAGGGCCAGTTCGAA 2073

Qy    244 ArgGluLeuGlnGluGluAspAspAspAspAspAspLeuGluAspGluVValuaspvalproarg 263
              ::::::||||||
Db   2074 AGGGATCTCCAAGCCCCGGGACGGAGCGAGACAGAATGAGGAG----------------AAG 2112

Qy    264 ArgArgSerSerAsp----------GlyGluProGlnSerGluAlaGluHisGlnArgArg 280
              |||||||           :::||||||
Db   2113 AGGAGGCAACTGCAGAGAACAGCTTCACGAGTATGAGACCGGAACCTGGAAGACAGAGCGAAAC 2172

Qy    281 AspLeuAlaArgGlyLeuLysSerSerProArgGlyLeuGluLeuGlnLeuGlnAsnAlaGlu 300
              |||                  :::||||||               |||::::||||||
Db   2173 GAACGTGCCCTGGCAGCTGCAGCAAGAAAGAAGACCTGGAAGGGGACCTGAAAGACCTGAG 2232

Qy    301 ValGlnSerLeuLeuSerTyrGlnArgMetArgAspSerProLeuSerLysSerArgArgPro 320
              :::||||||||   |||::: |||::: ::::|||::
Db   2233 CTTCAGGGCC---GACTCTGCCATCAAGGGAGGAGGAAGCCATCAAGCCAGCTACGCAAA 2289

Qy    321 Leu------AlaMetAsnAsp-----------------GluAspGluSerAlaPheArg--- 332
              :::       ||| ||| |||::||
Db   2290 CTGCAGGCTCAGATGAAGGACTTTCAAAGAGAGCTGAAGATGCCCGTGCCTCCAGAGAT 2349

Qy    333 AlaMetGluAlaArgAlaLysLeuAspGlnLysSerGlnLeuLysSerGlnLeuValLeuHisGly 352
              |||         |||||||
Db   2350 GAGATCTTTGCCACAGCCAAAGAAGAGAATGAGAAGAAAGCCAAGAGCTTG--------- 2397

Qy    353 PheGlyGluSerAspAspAspAspAspGluGluAspGluGluAspProSerGlu 372
              |||:::||| :::  :::|||||||
Db   2398 -----GAAGCAGACCCTCATGCAGCTCTACAAGAGGAC---------CTCGCCGCCGCTGAG 2442

Qy    373 AsnSerPheArgArgAlaProLeu---------ArgLeuSerGlyPheValGluLys 389
              |||                   |||||||||||
Db   2443 AGGGCTCGCCAAACAAGCGGACCTTCCAGAAGGAGAACTGGCAGGAGAGCTGGCCAGTAGC 2502

Qy    390 LeuLysSerAsnAspGluLeuLysSerAlaLeuAspArgIleLysThrTyrArgValAspAsp 409
              |||::              :::|||||||           |||:::||||||
Db   2503 CTGTCGGGAAGGAACGAACCACTCCAGGACAGAGAAGCCGCCTGAGCCCGGATCGCCCAG 2562

Qy    410 ValGluLysTyrLeuAlaProLysProMetGluPheGluAsnProLysProGlyProGlyTyr 429
              ::::|||::: :::  |||   ::::|||||||||
Db   2563 CTGGAGGAGGAGGTGGAGGAGCAGCAGG-CAACATGAGGAGGCCATGAGGAGCGACCGGGGTCC- 2620

Qy    430 PheAlaProArgLysIleProThr-----------------------------Arg 438
              |||||||||||   ::::
Db   2621 GCAAAGCCACACAGCAGGCCGAGCAGCTCAGCAACGAGCTGGCCAACGAGCCGCAGCACGG 2680

Qy    439 ProArgLysMet---LeuProLeuLeuIleGlySerAspProLysValGlnGluGluIle 457
              |||||||:::|||   ::::|||
Db   2681 CCCAGAAGAAATGAGAGTGCCC---------GGCAGCAGCTCGAGCGGCAGAACAAGGAGC 2731

Qy    458 ArgArgHisProSerThrGluTrpLysIleAlaLysGluAspValLeuThrAsnLeu 477
              |||||||:::|||  ::::|||
Db   2732 TCCGGGACAAGCTCCACGAGATGGAGGCGGAGGGG-GCCGTCAAGTCC----------- 2772
```

THIS PAGE BLANK (USPTO)

```
                Copyright    GenCore version 5.1.6
                Copyright (c) 1993 - 2005  Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on:        July 18, 2005, 12:25:31 ; Search time 5597 Seconds
                                           (without alignments)
                                           4372.936 Million cell updates/sec

Title:         US-10-736-868-2
Perfect score: 3291
Sequence:      1 MILFLFLLLGFCIAPLSA.........SKTKRFVGGNGAFDMPALGL 643

Scoring table: BLOSUM62
               Xgapop 10.0 , Xgapext   0.5
               Ygapop 10.0 , Ygapext   0.5
               Fgapop  6.0 , Fgapext   7.0
               Delop   6.0 , Delext    7.0

Searched:      34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters:     68479088

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=x1h
-Q=/cgn2_1/USPTO_spool/US10736868/runat_14072005_105510_2290/app_query.fasta_1.839
-DB=EST -QFMT=fastap -SUFFIX=rst -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10736868 @CGN_1_1_3437 @runat_14072005_105510_2290 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=10 -THREADS=10 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :     EST:*
               1: gb_est1:*
               2: gb_est2:*
               3: gb_htc:*
               4: gb_est3:*
               5: gb_est4:*
               6: gb_est5:*
               7: gb_est6:*
               8: gb_gss1:*
               9: gb_gss2:*
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	DB	ID	Description
c 1	484	14.7	300	1	AV178298	AV178298
2	349	10.6	476	4	BI741875	BI741875 kt82b03.y
3	167	5.1	3699	9	AY421012	AY421012 Pan trogl
4	162.5	4.9	5883	9	AY420735	AY420735 Homo sapi
5	162.5	4.9	5883	9	AY420737	AY420737 Mus muscu
6	159	4.8	4026	9	AY421011	AY421011 Homo sapi
7	155.5	4.7	2993	3	AK030316	AK030316 Mus muscu
8	153.5	4.7	7139	3	BC050866	BC050866 Mus muscu
9	153.5	4.7	7139	3	BC050868	BC050868 Mus muscu
10	152.5	4.6	3918	9	AY421013	AY421013 Mus muscu
11	152	4.6	2934	9	AY414528	AY414528 Homo sapi
12	151	4.6	3012	9	CL960352	CL960352 OsIFCC004
13	150.5	4.6	1128	3	CR679256	CR679256 Tetraodon
14	150.5	4.6	3431	5	AK030449	AK030449 Mus muscu
15	150	4.6	837	5	BW470454	BW470454 Mus muscu
c 16	150	4.6	2740	9	AK029960	AK029960 Mus muscu
17	150	4.5	2928	3	AY414530	AY414530 Mus muscu
18	148	4.5	3636.	3	AK029997	AK029997 Mus muscu
19	148	4.5	4613	3	AK033263	AK033263 Mus muscu
20	147.5	4.5	647	3	BX056132	CNS09FH4 Single re
21	147.5	4.5	6275	3	AK029236	AK029236 Mus muscu
22	146.5	4.5	3006	9	AY409299	AY409299 Homo sapi
23	146.5	4.5	3803	3	AL133606	HSM801462 Homo sapi
24	146.5	4.5	4277	3	CR627431	CR627431 Homo sapi
25	146	4.4	5504	9	AY407955	AY407955 Homo sapi
26	145.5	4.4	685	5	BW034508	BW034508
27	145.5	4.4	715	5	BW444409	BW444409
28	144.5	4.4	1168	6	CD506303	CD506303 CDA78-E05
29	144.5	4.4	4064	3	AK033354	AK033354 Mus muscu
30	144.5	4.4	5495	9	AY407957	AY407957 Mus muscu
31	144	4.4	1237	5	BU510057	BU510057 AGENCOURT
32	144	4.4	4269	3	AK038318	AK038318 Mus muscu
33	144	4.4	6294	9	AY410631	AY410631 Homo sapi
34	143.5	4.4	3191	3	CR623413	CR623413 full-leng
35	143	4.3	999	4	BM477724	BM477724 AGENCOURT
36	143	4.3	3395	3	BC057135	BC057135 Mus muscu
37	142	4.3	3914	3	BC032177	BC032177 Mus muscu
c 38	141.5	4.3	880	9	CG809737	CG809737 FSAAF38TF
39	141.5	4.3	3006	9	AY409300	AY409300 Pan trogl
40	141.5	4.3	3981	9	CL964889	CL964889 OsIFCC01
41	141	4.3	708	7	CK659858	CK659858 LP17954.5
42	141	4.3	771	7	CK478311	CK478311 AGENCOURT
43	141	4.3	2358	3	AK050669	AK050669 Mus muscu
44	141	4.3	3286	9	AK030978	AK030978 Mus muscu
45	141	4.3	4400	3	BC053625	BC053625 Homo sapi

ALIGNMENTS

```
RESULT 1
AV178298/c
LOCUS       AV178298              300 bp    mRNA       linear    EST 21-JUL-1999
DEFINITION  AV178298 Yuji Kohara unpublished cDNA:Strain N2 hermaphrodite
            embryo Caenorhabditis elegans cDNA clone yk563c9 3', mRNA sequence.
ACCESSION   AV178298
VERSION     AV178298.1  GI:5558199
KEYWORDS    EST.
SOURCE      Caenorhabditis elegans
  ORGANISM  Caenorhabditis elegans
            Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
            Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE   1  (bases 1 to 300)
  AUTHORS   Kohara,Y., Shin-i,T., Thierry-Mieg,J., Thierry-Mieg,D., Mitsuki,H.,
            Nishigaki,A., Motohashi,T., Zeng,Q., Watanabe,H., Sugimoto,A.,
            Sano,M., Miyata,A., Mitani,Y., Iida,K., Uesugi,H., Sugiyama,Y. and
            Nomoto,H.
  TITLE     Expressed genes in C.elegans
  JOURNAL   Unpublished (1999)
  COMMENT   Contact: Yuji Kohara
            Genome Biology Lab.
            National Institute of Genetics
            Yata 1111, Mishima, Shizuoka 411, Japan
            Tel: 81-559-81-6854
            Fax: 81-559-81-6855
            Email: ykohara@lab.nig.ac.jp.
FEATURES             Location/Qualifiers
     source          1..300
                     /organism="Caenorhabditis elegans"
                     /mol_type="mRNA"
                     /strain="N2"
                     /db_xref="taxon:6239"
                     /clone="yk563c9"
```

```
                                              /sex="hermaphrodite"                                FEATURES             Location/Qualifiers
                                              /dev_stage="embryo"                                    source           1..476
                                              /clone_lib="Yuji Kohara unpublished cDNA:Strain N2                     /organism="Strongyloides ratti"
                                               hermaphrodite embryo"                                                 /mol_type="mRNA"
ORIGIN                                                                                                               /db_xref="taxon:34506"
                                                                                                                     /dev_stage="l1"
Alignment Scores:                                                                                                    /lab_host="DH10B"
Pred. No.:                8.99e-38        Length:             300                                                     /clone_lib="Strongyloides ratti L1 pAMP1 v3 Chiapelli
Score:                     484.00        Matches:             88                                                      McCarter"
Percent Similarity:        97.78%        Conservative:         0                                                     /note="Vector: pAMP1 (Gibco); The library was constructed
Best Local Similarity:     97.78%        Mismatches:           2                                                      by Brandi Chiapelli and Dr. James McCarter at Washington
Query Match:               14.71%        Indels:               0                                                      University, St. Louis. The cDNA was made by using
DB:                              1       Gaps:                 0                                                      Dynabead oligo-dT priming (Dynal). PCR based library
                                                                                                                     using a modified protocol from the SMART PCR cDNA
US-10-736-868-2 (1-643) x AV178298 (1-300)                                                                           Synthesis Kit from Clontech. Directionally cloned into the
                                                                                                                     UDG sites of pAMP1. Nematodes were provided by Dr. Mark
Qy   554 GlyArgHisThrArgLeuArgTrpThrGlyAlaAsnGluLysGluIleProGlyLeuGly 573                                             Viney of Bristol, UK."
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   298 GGAAGGCATACCAGATTGAGATGGACTGGAGCAAATGAAAAAGAAATTCCAGGACTTGGA 239              ORIGIN

Qy   574 SerArgPheIleLeuProSerLeuAspProThrMetProAlaLeuAsnThrAlaPheSer 593              Alignment Scores:
         ||||||||||||||| ||||||||| |||||||||||||||| |||||||||||||||||              Pred. No.:                4.02e-24         Length:             476
Db   238 AGTCGCTTCATTCTCCCATNTCTTGATCCAACTATGCCANCCTTGAACACGCGCTTTCTCG 179              Score:                     349.00         Matches:             76
                                                                                    Percent Similarity:        60.12%         Conservative:        22
Qy   594 ThrGlnGlyArgAlaArgAspGluTrpAspThrMetPheLysIleProAsnAsnTrpAsn 613              Best Local Similarity:     46.63%         Mismatches:          51
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||              Query Match:               10.60%         Indels:              14
Db   178 ACTCAGGGGCGAGCCCGTGACGAGTGGGATACCATGTTCAAAATCCCGAATAACTGGAAT 119              DB:                              4       Gaps:                 4

Qy   614 ProGlyAspGluValGlyPheLysMetAsnSerLysThrLysArgPheValGlyGlyAsn 633              US-10-736-868-2 (1-643) x BI741875 (1-476)
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   118 CCTGGAGATGAAGTTGGGTTCAAATGAACTCAAAATGAACCAAACGATTCGTTCGGAGGAAAT 59           Qy   483 LeuAlaAlaLeuPheMetAspAspLysLeuGluAsnThrLeuLysGlyArgGlnMetLeu 502
                                                                                            :::                ||| :::             |||||||||||:||
Qy   634 GlyAlaPheAspMetProAlaLeuGlyLeu 643                                          Db     3 TTAGTTGCTTTATTGAAGGAACTGATCTGGCTGAAAAACTTGATAAACCATCATTTATG 62
         ||||||||||||||||||||||||||||
Db    58 GGAGCATTTGATATGCCTGCACTGGGATTG 29                                           Qy   503 ThrAspGluGlnLysGly------ArgThrArgValLysThrIleArgAlaLeuProArg 520
                                                                                            |||:||||:|||||                   |||  |||||||||||||||||||||
                                                                                    Db    63 TCAAATGAACAAAGAGTTTTAGTGTACACGTGGAAATGCAATTACAACATATCCACGT 122
RESULT 2
BI741875                                                                            Qy   521 LeuPheGlyAlaProThrAlaLysAlaGlu--------MetIleAspAlaLysValPheGln 538
LOCUS      BI741875        476 bp    mRNA      linear    EST 25-SEP-2001                     ::::|                |||                 ::: ||||||||||||||||||
DEFINITION kt82b03.y1 Strongyloides ratti L1 pAMP1 v3 Chiapelli McCarter            Db   123 ATGTATGCTGCTAAATTATTTGGTGATGAACCAAATTTTATGATAATAAAACAAAACAA 182
           Strongyloides ratti cDNA 5' similar to TR:P91120 P91120 SIMILAR TO
           HMG-BOX SINGLE-STRANDED DNA AND RNA BINDING PROTEINS. [1] /, mRNA        Qy   539 AspIleGluGlyArgProIleProLeuPhePheGluProLysGlyArgHisThrArg 558
           sequence.                                                                         :::|                 |||                 ||| ||||||||||||||||
ACCESSION  BI741875                                                                 Db   183 ATTGTTGAGGAAAGAGAAATTCCTCCACTTTTTTTGTACCAAAAGGAAAACATACAAGA 242
VERSION    BI741875.1     GI:15742831
KEYWORDS   EST.                                                                     Qy   559 LeuArgTrpThrGlyAlaAsnGluLysGluIleProGlyLeuGlySerArgPheIleLeu 578
SOURCE     Strongyloides ratti                                                               |||||||            ::::||||           |||||||||:|||:|::||
  ORGANISM Strongyloides ratti                                                      Db   243 TTAAGGTGGGTAACAGCAACAGAACAAGAAAATTCCCGGAATTGGAAGTAGAGATTTATTATA 302
           Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
           Panagrolaimoidea; Strongyloididae; Strongyloides.                       Qy   579 ProSerLeuAspProThrMetProAlaLeuAsnThrAlaPheSerThrGlnGlyArgAla 598
REFERENCE  1 (bases 1 to 476)                                                                |||::||                  |||              :::
AUTHORS    McCarter,J., Clifton,S., Chiapelli,B., Pape,D., Martin,J.,              Db   303 CCATCACTTGATCCAACTAGGCCAGCAATTAATTCAGTTGTATCAACAACAGGAAAGAA 362
           Wylie,T., Dante,M., Marra,M., Hillier,L., Kucaba,T., Theising,B.,
           Bowers,Y., Gibbons,M., Ritter,E., Bennett,J., Franklin,C.,              Qy   599 ArgAspGluTrpAspThrMetPheLysIleProAsnAsnTrpAsnProGly--------- 615
           Tsagareishvili,R., Ronko,I., Kennedy,S., Maguire,L., Beck,C.,                     |::|:||             ||:|||              :::
           Underwood,K., Steptoe,M., Allen,M., Person,B., Swaller,T.,              Db   363 AGAAATGAATATGAAACAACTTGGAAAAATACCAAATGCTTGGAAATCTGGAATATTTTG 422
           Harvey,N., Schurk,R., Kohn,S., Shin,T., Jackson,Y., Cardenas,M.,
           McCann,R., Waterston,R. and Wilson,R.                                    Qy   616 ------AspGluValGlyPheLysMetAsnSerLysThrLysArgPheValGlyGlyAsn 633
TITLE      The Washington Univ. Nematode EST Project, 1999                                                        |||:|:|||
JOURNAL    Unpublished (1999)                                                       Db   423 GAATTCGATGATACAAATCTGAAAATTAAT----------------TGGTGGGGACGGAAT 467
COMMENT    Contact: McCarter JP
           The Washington Univ. Nematode EST Project                                Qy   634 GlyAlaPhe 636
           Washington University School of Medicine                                          |||
           4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA            Db   468 AGTAAATTT 476  .
           Tel: 314 286 1800
           Fax: 314 286 1810
           Email: est@watson.wustl.edu                                              RESULT 3
           The library was constructed by Brandi Chiapelli and Dr. James           AY421012
           McCarter (bchiapel@watson.wustl.edu & jmccarte@watson.wustl.edu) at      LOCUS       AY421012              3699 bp    DNA      linear    GSS 17-DEC-2003
           Washington University, St. Louis. DNA Sequencing by: Washington          DEFINITION  Pan troglodytes ACINUS gene, VIRTUAL TRANSCRIPT, partial sequence,
           University Genome Sequencing Center St. Louis.                                       genomic survey sequence.
           Seq primer: -40RP from Gibco                                             ACCESSION   AY421012
           High quality sequence stop: 306.                                         VERSION     AY421012.1   GI:39776969
                                                                                    KEYWORDS    GSS.
```

```
SOURCE      Pan troglodytes (chimpanzee)
ORGANISM    Pan troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE   1 (bases 1 to 3699)
AUTHORS     Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejariwal,A.,
            Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
            Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J.,
            Adams,M.D. and Cargill,M.
TITLE       Inferring nonneutral evolution from human-chimp-mouse orthologous
            gene trios
JOURNAL     Science 302 (5652), 1960-1963 (2003)
PUBMED      14671302
REFERENCE   2 (bases 1 to 3699)
AUTHORS     Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejariwal,A.,
            Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
            Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J.,
            Adams,M.D. and Cargill,M.
TITLE       Direct Submission
JOURNAL     Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,
            Rockville, MD 20850, USA
COMMENT     This sequence was made by sequencing genomic exons and ordering them
            based on alignment.
FEATURES             Location/Qualifiers
     source          1..3699
                     /organism="Pan troglodytes"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9598"
     gene            <1..>3699
                     /gene="ACINUS"
                     /locus_tag="HCM7415"

ORIGIN

Alignment Scores:
Pred. No.:          8.34e-05        Length:          3699
Score:              167.00          Matches:         115
Percent Similarity: 36.79%          Conservative:    66
Best Local Similarity: 23.37%       Mismatches:      176
Query Match:        5.07%           Indels:          136
DB:                 9               Gaps:            23
```

US-10-736-868-2 (1-643) x AY421012 (1-3699)

```
Qy     23 ProSerThrSerAspAlaProGlyAlaLeuLeuSer---SerLeuValGlySerHis 41
             |||||         ||       ||||     ::: |||||||
Db    157 CCGAGTACCTCACGCAAGATGCCGGAGCTGGAGGAGGTGACTCTGGACGGGAAGCCTCTT 216

Qy     42 GlnLysLeuProLeuAlaProSerMetGluAlaLeuGlyLeuMetGluValGlnPheVal 61
             ||||       |||        |||||||||||
Db    217 CAGGGCGTGCGGGTGACCGACCGAGCCGACCACTGGAGCAGCGAGGC--------- 264

Qy     62 AspAlaLeuIleLeuTyrLysGlyGln-------MetGluMetAlaLysGly----- 75
             |||       ||| |||||       ::::::
Db    265 -----CTAGCCAGAGGGGCCGAGAAGTGCCTGGTGCCTGTGTCCAAGGGGCCTCTA 318

Qy     76 --------------------------------AlaPheLysThrGlnLeuGlu 82
                                                |||||||||::::
Db    319 ATGCTAGAAAAATTTACAGAAGAAACACTCAACACACCCCATGCTGCATTCCAGCCAAATTCCCAG 378

Qy     83 ValLeuGluLysValHisProAspGlnPheAspLysTyrLysLysLeuLysValAspAsp 102
             :::|||||:::      |||            ||| |||           ::: :::
Db    379 ATTGGTGGAGAAGAATGAGCCAGGAACAGTTTCATAAAACAGTATCTGGAAAAGCAGCAGGAG 438

Qy    103 LeuAlaAlaAspAlaValMetGlnGlnAla---GluMetAlaLysGlnProLysSer 121
             |||     |||           ||| |||||  :::|||||:::||::
Db    439 CTACTTAGGCAGCGTCTGGAACGTGAAGCTCGAGAAGCTCGAGAAGAAGCTTCA 498

Qy    122 GlyAsnAlaPheLeuAspMetLeuAsnGlyAsnGlyIle-----------------Pro 135
             :::                |||             :::
Db    499 GCTGAGTCGACGGAGGACGAGATGATCCATCCTGAGGGAGTGGGCTTCCCGTGCTGCTCCTGAC 558

Qy    136 IleGlySerSerIleArgGlyLeuGluAspAlaIleArgThrGlnArgAspMetGluAsn 155
             |||||::|::                       |||            ::: :::::::
```

```
Db    559 TTTCAGAGCAGCCTGGAGACCACCAGACTCGAGCTCAGCAGACATTCGCCAGAAAAAGC 618
Qy    156 ThrAspProSerGluGlnIleAlaLysValMetAspLysPheGlnIleLeu 175
              :::                    |||||:::
Db    619 TCCTCAATTCTGAAGAG------------- 636
Qy    176 ProGlyLeuValAlaAsnMetIleAlaGlyLysAsnProPheMetProGlnGlnMet 195
             |||                        ::: :::
Db    637 -------------------AAAGGTGGACTCTGATGATGAGAAACCA--- 663
Qy    196 ArgLysAlaGlnIleAlaAlaProSerValPheGlnGlnAlaGlnAlaLeuAlaArgAlaMet 215
             |||||        |||                       ::::::::::::
Db    664 AGGAAAGGAGAAAGACGATCATCTAGGGTC---AGACAGCGCAAGAGCAGCTAAACTGTCT 720
Qy    216 LeuGlyLysAsnAlaProValAlaGlyAlaGlyArgGly-----GlnArgMetGlu 233
             |||               |||                    :::|||
Db    721 GAGGGC--AGCCAACCTGCTGA-GGAGGAAGAGGATCAAGAAGAACACCTTCCAGAAACCT 776
Qy    233 tMetAsnArgValAspGlnArgMet------GlnGlnArgGluLeuGlnGluGluAspGl 251
               : |||              :::  ::::::::||||||::::::
Db    777 AAGGGTCAGGACGATCGAAATTTGAAAACAGAGGAAGAGGAGGAAGAGGAGGAGGAGGA 836
Qy    251 uAspAspAspAspLeuGluAspGluAspSerSerAspSerGlyGluPr 271
             ||::::||:::::::                                 |||
Db    837 GGAAGAAGATGATGAAGAGGAGGAAGAT 875
Qy    271 oGlnSerGluAlaGluHisGlnArgArgAspLeuAlaArgArgLeuLysSerSerProAr 291
             |||                                                        |||
Db    876 ACAAAAATCTAGGGAG-----------GCACCAAT 899
Qy    291 gLeuLysGlyGluLeuGlnAsnAlaGluValGlnSerLeuLeuSerTyrGlnArgMetAr 311
             |||||||||                         ::: :::
Db    900 CCTGAAAGAGTTTAAGGAAGAAGAGGGAAGATCCTAGAGTAAAACCAGAAGAGAAGATGAT 959
Qy    311 gAspSerProLeuSerLysArgArgProLeuAlaMetAsnAspGluAspGluSerAlaPh 331
             |||                                                ||:::||||
Db    960 GGAT-----------------GAGAGACCCAAAACAAGATCCCAGGAACAGCAGAGGTGTTNNN 1004
Qy    331 eArgAla----------------------MetGluAlaArgAlaLeuAsnLysAspGlnLysSe 344
             |||                         |||||||||
Db   1005 GAGAGGAGGAGGAGATTTCAAGATCCCAGGAAGAGGCTAGA----------AAAAG 1049
Qy    344 rGlnLeuValLeuGlyLeuHisGlyPheGlyGluSerAspAspAspGluAspGluGluAs 364
             |||||       :::                  |||
Db   1050 TCATCTGGCC-----------------AGACAGCAGGAGGAAGGA 1079
Qy    364 pGluAsnLeuIleAspProSerGluAsnSerPheArgArgAlaProLeuArgLeuSerSe 384
             |||              ||:::               :::
Db   1080 AATGAAAACAACATCTCCCCTGAGGAGAAGAGAGAA------ATAAAATCTTCACN 1133
Qy    384 rGlyPheValGluLysLeuLeuLySer 397
             |||           ||||||
Db   1134 NGGCTTAAAGGAAAAATCGAAGTCTCCTTCCCCTGCGACTGAGATCGAAAGAA 1193
Qy    397 sSerAlaLeuAspArgIleLysTyrArgValAspAspValLysTyrLeuLysAlaProLy 417
             |||          :::                           ::::::
Db   1194 GACCTCACTTGTAGCGCTGCCGCAGACAAACTGCCAGCGAGGAG-----ACTCCTCC 1247
Qy    417 sProMet---------------------GluPheAsnProLysProGlyTyrPheAlaProAr 433
             |||                         :::
Db   1248 ACCTTTACTAACAAAGGAGCAATCTTCACCACCACCTCCATCCAGCTCCATAGCGAAGA 1307
Qy    433 gLysIleProThrArgProArgLysMetLeuProLeuLeuIleGlySerAspProLysVa 453
             |||                                          ||| 
Db   1308 AGAAATAGAGCCCATGGAAGGCCCAGCCCCCCCTGTCCTCATTCAGTTATCTCCT------ 1362
Qy    453 lGlnGluGluIleArgArgHisProSerThrGlu 464
             |||||::::::::::::
Db   1363 --------------------------------CCTAATACAGAT 1374
```

RESULT 4
AY420735
LOCUS   AY420735   5883 bp   DNA   linear   GSS 17-DEC-2003

```
DEFINITION   Homo sapiens MYH9 gene, VIRTUAL TRANSCRIPT, partial sequence,
             genomic survey sequence.
ACCESSION    AY420735
VERSION      AY420735.1  GI:39776692
KEYWORDS     GSS.
SOURCE       Homo sapiens (human)
  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1  (bases 1 to 5883)
  AUTHORS    Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejariwal,A.,
             Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
             Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J.,
             Adams,M.D. and Cargill,M.
  TITLE      Inferring nonneutral evolution from human-chimp-mouse orthologous
             gene trios
  JOURNAL    Science 302 (5652), 1960-1963 (2003)
   PUBMED    14671302
REFERENCE    2  (bases 1 to 5883)
  AUTHORS    Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejariwal,A.,
             Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
             Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J.,
             Adams,M.D. and Cargill,M.
  TITLE      Direct Submission
  JOURNAL    Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,
             Rockville, MD 20850, USA
COMMENT      This sequence as made by sequencing genomic exons and ordering them
             based on alignment.
FEATURES             Location/Qualifiers
     source          1..5883
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
     gene            <1..>5883
                     /gene="MYH9"
                     /locus_tag="HCM7323"
ORIGIN
```

Alignment Scores:

Pred. No.:	0.00046	Length:	5883
Score:	162.50	Matches:	136
Percent Similarity:	35.90%	Conservative:	88
Best Local Similarity:	21.79%	Mismatches:	213
Query Match:	4.94%	Indels:	192
DB:		Gaps:	9

```
US-10-736-868-2 (1-643) x AY420735 (1-5883)

Qy    12 GlyPheCysIleAlaProLeuSer-----AlaGlnSerProSerThrSerAsp------     27
            |||   ::|        |||     ||||||
Db  3860 GGCTTCTGCAGCCAGTCCGACAGCAAGTCTCAGCAGCTTCACCAAGGACTTCTCCGGCGCTGG 3919

Qy    28 AlaProGly------AlaLeuLeuSerSerLeuValGlyLysSerHisGlnLysLeuPro     45
            :::|||      ||      ||||||   ::::::   |||||
Db  3920 AGTCCCAGCTGGAGGACACTCCAGGAGCTG-CTGCAGGAGGAGAACCGGCAGAAGCTGAGC  3978

Qy    46 LeuAlaProSerMetGluValAlaLeuGlu-----------                       54
            ||||   :::::|   |||||
Db  3979 CTGAGCACCAAGCTCAAGCAGCTGGAGGACGTGAGGAGGACGAGAAGAATTCCTTCGGGGAGGAGGAG 4038

Qy    55 -------------------------------------LeuMetGlyValGlnPhe         60
                                                 |||   :::
Db  4039 GAGGAGGAGGAGGCCCAAGCCAGCAAGCTCCAGCAGGAGGAGCAGATCGCCCATGCCCAGGTG 4098

Qy    61 ValAspAlaLeuIleLysLysGlyMetGluMetAlaLysGlyAlaPheLysProSerGl     80
            ::::|||||||   :::|||||||     |||||   :::||||
Db  4099 GCCGAC------ATGAAAAGAGATGAGGACAGTGTGGGGGTTGCCTGGAAACTGCT      4149

Qy    81 LeuGluLeuGluLysGluValHisProAsp-----------                       90
            :::::|     ||||||||||
Db  4150 GAGGAGGTGAAGAGGAGAGAAGCTCCAGAAGGAGGAGGACCTGAGCCGCCACCGAGGAG    4209

Qy    91 ---------GlnPheAspLysTyrLysLysLeuLys-----ValAsp                101
                     :::||||||
Qy   102 AspLeuAlaLeuAlaAspAlaValMetGlnAlaGluMetAlaLysGlnProLysSer     121
            :::|||||    :::|||   |||
Db  4210 AAGGTGGCGCCCTACGACAAGCTGGAGAAGACCAAGAGACGCGGCGTCGAGCGAGCTGGAC   4269
            |||::::
Db  4270 GACCTGCTGGTGGACCACCAGCCGTGCAACCTGGAGAAGAAGCAG                    4329

Qy   122 GlyAsnAlaPheIleAspMetLeuAsnGlyAsnGlyIleProIleGly-----------    137
            |||
Db  4330 -----TTTGACCAGCTCCTGGCGGAGAGAGACCATCTCTGCCAAGTATGCA             4383

Qy   138 ------SerSer                                                    139
Db  4384 GAGGAGCGCGACCGGGCTGAGGCGGAGGCCCGAGAAGGAGGAGGACCAAGGCTCTGTCGCTG  4443

Qy   140 IleArgGlyLeuGluAspGlnAlaIleLeuArgThrGlnAsnThrGlnAspHisAsnThrAsnProProSer 159
            ||||:::::||   ::::::||
Db  4444 GCCCGGCGCCTGGAGGAAGCCATGGACGAAGAGCAGCAGCGCTGGAGCGGCTCAAC-----------      4497

Qy   160 GluGlnIleLeuAlaLysAlaValMetAspLysPheGlnIleThrGlnIleLeuProGlyLeuVal      179
            ::::||||:::   |||
Db  4498 -----AAGCAGTTCCGCACGGAG---------------------------------ATG         4518

Qy   180 AlaAsnMetIleAlaGlyLeuLysAsnProPhe---LysMetProGlnMetGlnMetLysLysAla  198
            ::::|||          |||||
Db  4519 GAGGAACCTTATGAGCTCCAAGGATGTGGGCAAGAGTGTCCACGAGTGGGAGAAGTC-          4577

Qy   199 GlnAlaAlaProSerSerValPheGlnGlnIleAlaLeuGlnIleArgAlaMetLeuGlyLys      218
            ||||||||   :::
Db  4578 CAAGCCGGCCCT---------------------------AGAGCACAGTGGGAGGA           4607

Qy   219 Asn----------AlaProValAlaGlyArgGlyGluGlnArgMetMetAsn            235
            :::   ||       |||
Db  4608 GATGAACAGCAGCTGGAGGACCAGCTCGAAGAGCTGCAGGCCACCGAAGATGCCAAGCT      4667

Qy   236 -ArgValAspGlnArgMetGln-----------------------GlnArgGluLeuGluGl 249
            :::|||::||                                   |||||
Db  4668 GCGGGTTCGGGAGGTTCAACCTGCCAGGCCATGAAGGCCCAGTTCGAGCGCGGACCTGCAGGGCCG 4727

Qy   249 uAspValAspProAspAspLeuAspLeuGluAspValProProArgArgSerSerArgAspGl 269
            :::   :::   :::   |||
Db  4728 GGACGAGCGAGAGCAGGAGAGAAGCAGCACGCTGGTCAGACAGGTGCGGGACGATGGAGGC     4787

Qy   269 yGluProGlnSerGluValAlaGluHisGlnIleAsnArgAspAspArgArgLeuLysSerSerGl 289
            :::   :::            :::
Db  4788 AGAGCAGCACGGTGGAGAGGAAGCAGCAGCGCCGCCGAGAGATCGCAGTGGCCGCCCGGAAGAAGCTGGA 4847

Qy   289 rProArgLeuLysGluLeu--------------                               295
            :::||||||
Db  4848 GATGGAACCTGAAGAGCCCTGGAGGCTCACGTGGAGCCCAACAGAGAACCGGACGAAGAGC     4907

Qy   296 --------------LeuGlnAsnAlaGluValGlnSerLeuLeuSerGlnLysAsnLeuIleAspSerArgAs 312
                          :::   :::                ::: |||
Db  4908 CATCAAACAGTTGCCAGAGCTGTCCAGCGCCGATGAAGGACACTCGGCGCTGGAGCTGGATGA       4967

Qy   312 pSerProLeuSerArgArgProLeuAlaMetAsnAspAspGluSerAlaPheAr           332
            ||||||
Db  4968 CACCCGGCCGCTTCGGTGGAGGATCCTGGCCCCAGGCCCAAAGACAGAAGAAGCTGAA     5027

Qy   332 gAlaMetGluAlaLysArgAlaLysAspLeuLysAspLysGlnLysSerGlnLysValLeuGlnLeuHisGl 352
            :::||||                              --------
Db  5028 GAGATGATCCAGTTGCAGGAGGAACTGGCCAGC-------------------GAGCATGGAGGC          5072

Qy   352 yPheGlyGluSerArgAspAspAspAspGluAspGluLeuGluLeuAspProSerGl            372
            :::                                    :::
Db  5073 CGCGGAGCCTGCCCAAGGCCCAGCCCCAGGAGCCGGATAGAGCTGCGACGAGATCGC            5132

Qy   372 uAsnSerPheArgArgAlaProLeuArgLeuSer                              383
            :::   ||||
Db  5133 CAACAGCAGCCGCAAAGGACCCTGGCTTAGGAGGAGAAGCGCGTCTGGAGGCCCCGAT     5192

Qy   384 -SerGlyPheValGluLysLysLeuLysSer                                 395
            :::   :::||||||
```

```
Score:                  162.50              Matches:        124
Percent Similarity:      34.07%             Conservative:    92
Best Local Similarity:   19.56%             Mismatches:     196
Query Match:              4.94%             Indels:         225
DB:                                         Gaps:            23

US-10-736-868-2 (1-643) x AY420737 (1-5883)

Qy   39 LysSerHisGlnLysLeuProLeuAlaProSerMetGluAlaLeuGlu------ 54
            :::|||  ||||||| ||||||::::: :::|||
Db 3958 GAGACCGGCAGAAGCTGAGCACCAAGCTCAAGCAGATGGAGGATGAGAAAAAC 4017

Db   54 ------------------------------------------- 54
Db 4018 TCCTTCAGGGAGCGACGACGAGGAGGAGGACGCCAACTTGGAGGAGCGATC 4077

Qy   55 --LeuMetGlyValGlnPheValAspAspAlaLeuGluPheLysGlyGlnMetGluMetAla 73
           :::            :::                        :::
Db 4078 GCCACGCTCCATGCCCAGGTGACCGAC------ATGAGGAAGAGATGGAGGACGCGGT 4128

Qy   74 LysGlyAlaPheLysThrGlnLeuGluValLeuGluGluValHisProAsp------ 90
           |||                                          :::||||  |||
Db 4129 GTAGGGTGCCTGGAGACTGCAGAGGAGGCGGAAGCCGGAGGCTTCAGAAGGACTTGGAAGGC 4188

Qy   91 ------------GlnPheAspLysTyrLysLysLeuLys-- 99
                               :::||||||| :::|||  |||
Db 4189 CTGAGCCAGCCAGGCTTCTGCCGCTACGATAAGCTGGAGAAGAACCAAGACA 4248

Qy  100 ------ValAspLeuAlaAlaAspAlaValMetGlnGlnLeuHisGluMet 114
                  :::|||||||
Db 4249 CGGCTGCAGCAGGAGCTGGACGACCTGGTTGACCTGGACAGGAGGGCCAGAGCGTC 4308

Qy  115 AlaLysLeuGlnProLysSerGlyAsnAlaPheIleAspMetLeuAsnGlyAsnGlyIle 134
           :::  ||||||:::
Db 4309 TCCAACCTGGAAAAGAACAGACAGAGAAG------TTCGACCAGCTCCTAGCCGAGGAGAAG 4362

Qy  135 ProIleGly--- 137
           |||
Db 4363 ACCATCTCGCCAAGTATGCCAGAGAGCTGACCGAGGCTGAGGCTGGCCGTGAGAAG 4422

Qy  138 ------SerSerIleArgGlyLeuGluAspAlaIleAlaArgThrGlnArgAsp 152
                   :::            :::
Db 4423 GAGACAAAGGCGCTATCACTGGCCCGGGCCTTGGAGGAGGCCATGGAGCAGAAGGCAGAG 4482

Qy  153 MetGluAsnThrAspProSerGluGlnIleIleAlaAlaValMetAspLysProPheGlnThr 172
           :::|||                                          :::
Db 4483 CTGGAGCGGCTCAAC----------AAGCAGTTCCGACACG 4512

Qy  173 GlnIleLeuProGlyLeuValAlaAlaAsnMetIleAlaGlyLysAsnProPheLysMetPro 192
           |||||||  :::            :::|||:::
Db 4513 GAGATGAGGACCTCATGAGCTCCAAGATGACGTGGCAAGAGTGTCCAAGACTGGAG 4572

Qy  193 GlnGlnMetArgLysAlaGlnAlaAlaProSerValPheGlnGlnAlaLeuAlaGln 212
           :::                                          ||||||
Db 4573 AAGTCCAAGCGG----------GCCTTGGAGCAG 4596

Qy  213 ArgAlaMetLeuGlyLysAsnAlaProValAlaAlaGlyArgGlyGluGlnGlnArgMet 232
           :::                                          |||||
Db 4597 CAGTG------GAGAGATGAAGACC 4617

Qy  233 MetMetAsnArgValAspGlnMetGln--- 242
           :::            :::||||||:::
Db 4618 CAGCTGGAGGAGCTGGAGGATGAGCTGGAGCCAGCTGCCACGGAGATGCTCCGGGAG 4677

Qy  243 ------GlnArgGluLeuGlnGluGluLeuGluAspAsp 252
                  |||:::|||||:::
Db 4678 GTGAACCTGCAGGCCATGAAGGCCCAGTTTGAGCGGGACCGGGATGAACAG 4737

Qy  253 AspAspLeuLysAspGluValAspGluAspArgArgSerSerAspGlyGluProGln 272
           :::  ||| :::                                    :::
Db 4738 AGCGAGGATGAAGAGAGCAGCTGGTCAGACTGCGGGAGATGCGGGAGCTGGAG 4797

Qy  273 SerGluAlaGluHisGlnIleArgArgAspLeuGluLeuLysSerSerProArgLeu 292


Db 5193 CGCCCAGCTGGAGGAGGAGCTGGAGGAGGAGCCAACACGGAGCTGATCAACGACCG 5252
Qy  395 uLeuLysSerAlaLeuAspArgIleLysTyrArgValAspAspValGluLysTyrLeuAl 415
           |||||  |||            :::|||||||  :::
Db 5253 GCTGAAGAAGGCC----------AACCTGCGAGATCGACCAGATC---- 5286

Qy  415 aProLysProMetGluPheAsnProLysProGlyTyrPheAlaProArgLysIl 435
           |||          :::|||||||
Db 5287 ------------------AACAC-CGACCGAACCCTGA- 5306

Qy  435 eProThrArgProArgLysMetLeuProLeuLeuIleGlySerAspProLysValGlnGl 455
           :::||||||   |||||||:::
Db 5307 ---GCGCAGCCACGCCC-AGAGAACGAGAATGCTCCGCAGCCAGCTGGAACGCCAGAAC 5361

Qy  455 uGluIleuArgArgHisProSerThrGluTrpPheLysIleAlaLysGluSerArgValLeuTh 475
           |||            :::|||||||  :::
Db 5362 AAGGAGCTTAAGTCAAGCTGCCAGGAGATGG-----------AGGGCCACTGTC 5403

Qy  475 rAsnLeuLysAsnAsnProSerLeuAlaAlaLeu---------PheMetAspAs 490
           :::|||||||                  :::|||||||
Db 5404 AAGTCC--AAGTACAAAGGCCTCCATCACCGCCTCGAGGCCAAGATTGCACAGCTGGAGGA 5462

Qy  490 pLysLeuGluAsnThrLeuLysGlyTyrAsnThrAspGlnMetLeuThrArgLysGlyArgTh 510
           ::::||||  |||||||||:::
Db 5463 GCAGCTGGACAACGACCGAGACCGAAGGAGCGCCAGCCTGC--AAACAGGTGCGTCGGAC 5519

Qy  510 rArgValLys 513
           |||
Db 5520 CGAGAAGAAG 5529


RESULT 5
AY420737
LOCUS       AY420737               5883 bp    DNA     linear   GSS 17-DEC-2003
DEFINITION  Mus musculus MYH9 gene, VIRTUAL TRANSCRIPT, partial sequence.
            genomic survey sequence.
ACCESSION   AY420737
VERSION     AY420737.1  GI:39776694
KEYWORDS    GSS.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 5883)
  AUTHORS   Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejariwal,A.,
            Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
            Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J.,
            Adams,M.D. and Cargill,M.
  TITLE     Inferring nonneutral evolution from human-chimp-mouse orthologous
            gene trios
  JOURNAL   Science 302 (5652), 1960-1963 (2003)
   PUBMED   14671302
REFERENCE   2  (bases 1 to 5883)
  AUTHORS   Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejariwal,A.,
            Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
            Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J.,
            Adams,M.D. and Cargill,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,
            Rockville, MD 20850, USA
COMMENT     This sequence as made by sequencing genomic exons and ordering them
            based on alignment.
FEATURES             Location/Qualifiers
     source          1..5883
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
     gene            <1..>5883
                     /gene="MYH9"
                     /locus_tag="HCM7323"
ORIGIN                              Length:        5883
Alignment Scores:
Pred. No.:       0.00046
```

```
                                                                    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                                                                    Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
                                                  REFERENCE    1  (bases 1 to 4026)
                                                  AUTHORS    Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejariwal,A.,
                                                             Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
                                                             Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J.,
                                                             Adams,M.D. and Cargill,M.
                                                  TITLE    Inferring nonneutral evolution from human-chimp-mouse orthologous
                                                           gene trios
                                                  JOURNAL   Science 302 (5652), 1960-1963 (2003)
                                                  PUBMED   14671302
                                                  REFERENCE    2  (bases 1 to 4026)
                                                  AUTHORS    Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejariwal,A.,
                                                             Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
                                                             Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J.,
                                                             Adams,M.D. and Cargill,M.
                                                  TITLE    Direct Submission
                                                  JOURNAL   Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,
                                                             Rockville, MD 20850, USA
                                                  COMMENT   This sequence as made by sequencing genomic exons and ordering them
                                                             based on alignment.
                                                  FEATURES             Location/Qualifiers
                                                  source          1..4026
                                                                  /organism="Homo sapiens"
                                                                  /mol_type="genomic DNA"
                                                                  /db_xref="taxon:9606"
                                                  gene            <1..>4026
                                                                  /gene="ACINUS"
                                                                  /locus_tag="HCM7415"
                                                  ORIGIN
                                                  Alignment Scores:
                                                  Pred. No.:                 4026
                                                  Score:               159.00
                                                  Percent Similarity:    36.06%        Matches:         159
                                                  Best Local Similarity: 22.39%        Conservative:     97
                                                  Query Match:            4.83%        Mismatches:      260
                                                  DB:                                  Indels:          196
                                                                                       Gaps:             35

                                                  US-10-736-868-2 (1-643) x AY421011 (1-4026)
```

```
Db       4798  GATGAGGAGAAGCAGCCGCTCCATGGCCATGGGCCGCACGCAAGAAACTGGAGATGGATCTG  4857
                   |||               :::       |||       |||
Qy        293  LysGluLeu-------------------------------------------------------  295
                   |||:::|||
Db       4858  AAGGACCTGGAGGCACACATTGACACAGCCAATAAGAACCGGGAAGAGAGGCCATCAAACAG  4917

Qy        296  LeuGlnAsnAlaGluValGlnSerLeuLeuSerTyrArgMetArgAspSerProLeu  315
                   |||:::|||       :::       |||:::|||
Db       4918  CTGCGGAAGCTTCAGGCCCAGATGAAGGACTGCATGCGGGAGCTGACGACCACGCGCGCC  4977

Qy        316  SerLysArgArgProLeuAlaMetAsnAspGluSerAlaPheArgAlaMetGlu  335
                   |||||||       :::|||:::|||
Db       4978  TCCCGGGAGGAGATCCTCGGCCCAGGCCAAGGAGAGAATGAGAAGAGCATGGAG  5037

Qy        336  AlaArgAlaLysAlaLysSerGlnLysSerGlnLeuValLeuGlyLeuHisGlyPheGlyGlu  355
                   |||
Db       5038  GCC--------------------------------------------------------------  5082

Qy        356  SerAspAspGluAspGluGluAspGluAsnLeuIleAspProSerGluAsnSerPhe  375
                   :::|||       :::|||       |||       ||||||
Db       5083  GCTAAGGCGTCAGGCGACCAGGAACGGCAGCGTGGCTGATGAGATCGCCAACAGCAGT  5142

Qy        376  ArgArgAlaProLeuArgLeuSer--------------------------------SerGlyPhe  386
                   :::
Db       5143  GGCAAAGGGGCCCTAGCATTAGAGGAGAAGCCGGCAGCTGGAGGCCCGCATTGCCCAGCTG  5202

Qy        387  ValGluGluLysLysSer--------------------------------AsnAspGluLysLysSer  398
                   |||:::|||||||       ||||||
Db       5203  GAGGAGGAGCTGGAGGAGGAACAGGGAACACGGAGCTGATCAACGATCGGCTGAAGAAG  5262

Qy        399  AlaLeuAspArgArgIleLysTyrArgValAspAspAspProValGluLysTyrLeuAlaProLysProLysPro  418
                   ||||||       |||
Db       5263  GCC---------------------------------------------------ACAGAAGGAATG-AGA  5332

Qy        419  MetGluPheAsnProLysProGlnProGlyTyrPheAlaProArgLysIleProThrArg  438
                   :::|||       |||:::       ||||||
Db       5333  ATGCGGCAGACAGCAGC---TGGAACCGGACACAAGGAGGCTCAAGGCCCAAGCTGCAGGAAA  5389

Qy        439  ProArgLysProMetLeuProLeuLeuIleGlyLeuSerAspProLysValGlnGluIleGluArg  458
                   |||||||
Db       5390  --ArgValLeuThrAsnLeuLysAsnAsnProSerLeuAlaAlaLeu-------------------  486

Qy        459  ArgHisProSerThrGluTrpLysIleAlaAlaGluSer-------------------------------  471
                   |||:::|||       |||
Db       5449  GCACAGCTGGAGGAACAGCTGACACAGGAGCTGATCAACGATCGGCTGAAGAAG  5505

Qy        506  GlnLysGlyArgThrArgValLysThrIleArgLeuProArgLeuPheGlyAlaPro  525
                   |||       |||       :::
Db       5506  CAGGTGCGCCGGACAGAGAGAAGAGTCTGAAGACGTGTCGCTGCAGGTG---------  5553
```

```
                                                  RESULT 6
                                                  AY421011
                                                  LOCUS      AY421011         4026 bp    DNA    linear    GSS 17-DEC-2003
                                                  DEFINITION  Homo sapiens ACINUS gene, VIRTUAL TRANSCRIPT, partial sequence,
                                                             genomic survey sequence.
                                                  ACCESSION  AY421011
                                                  VERSION    AY421011.1  GI:39776968
                                                  KEYWORDS   GSS.
                                                  SOURCE     Homo sapiens (human)
                                                  ORGANISM   Homo sapiens
```

Qy 156 ThrAspProSerGluGlnIleAlaAlaLysAlaValMetAspLysPheGlnThrGlnIleLeu 175
              ::||||||||:::
Db 619 TCCTCAATTTCTGAAGAG   636

Qy 176 ProGlyLeuValAlaAsnMetIleAlaGlyLysAsnProPheLysMetProGlnGlnMet 195
Db 637 --------------------AAAGGTGACTCTGATGATGAGAAACCA   663

Qy 196 ArgLysAlaGlnAlaAlaProSerValPheGlnIleAlaLeuAlaAlaGlnArgAlaMet 215
              |||||||| ::: :::||||||||||| :::
Db 664 AGGAAAAGGAAGAAGACGATCATCTAGGGTC---AGACAGGCAAGAGCAGCTAAACTGTCT 720

Qy 216 LeuGlyLysAsnAlaProValAlaGlyGlyValGlyArgGly-----GlnArgMetMe 233
              ||| ||| :: ||
Db 721 GAGGGC---AGCCAACCTGCTGA-GGAGGAAGAGGATCAAGAAACACCTTCCAGAAACCT 776

Qy 233 tMetAsnArgValAspGlnArgMet------GlnGlnArgGluLeuGlnLeuGluAspGl 251
              :||:||||||:::|||| :: ::
Db 777 AAGGGTCAGACAGCAGATCGAAATTTGAAAACAGAGGAGAAGAGAGAGGAGGAGGAGGA 836

Qy 251 uAspAspAspAspLeuGluAspGluGluAspValProArgArgArgSerSerGlyGluPr 271
              :|::::|||||||| :::::|||
Db 837 GGAGAAGATGATGAAGAAGAGGAA------GGTGATGATGAGGG 875

Qy 271 oGlnSerGluAlaGluHisGlnAsnArgAspLeuAlaArgArgLeuLysLysSerProAr 291
              :|| |||
Db 876 ACAAAAATCTAGGGAG------GCACCAAT 899

Qy 291 gLeuLysGluLeuLeuGlnAsnAlaGluValGlnSerLeuLeuSerTyrGlnArgMetAr 311
              || |||
Db 900 CCTGAAAGAGTTTAAGGAAGAAGGGGGA------GAGATACCTAGAGTAAA 944

Qy 311 gAspSerProLeuSerLysArgArgProLeuAlaMetAsnAsnAspGluGlu------ 328
              ::: |||
Db 945 ACCAGAGGAGATGATGAGATCCAAAACAAGATCCCAGGAACAAGGAGGTGTTAGA 1004

Qy 329 ------SerAlaPheArgAlaMetGluAlaArgAlaLysArgAlaLeuAspGlnLysSerGlnLe 346
                   ||| ::: ||||||
Db 1005 GAGAGGAGGAGGATTACAAGATCCCAGGAAGAGGCTAGA-------AAAGTCATCT 1055

Qy 346 uValLeuGlyLeuHisGlyGlyPheGlyGluSerAspAspAspGluAspGluGluAspGluAs 366
                                                 :: ::: ||
Db 1056 GGCC-----------------------------------AGACAGCAGCAGGAGAAGAAATCAA 1085

Qy 366 nLeuIleAspProSerGluAsnSerPheArgArgAlaProLeuArgLeuSerSerGlyPh 386
              :||| ::: :::|||| :::::: :::
Db 1086 AACAACATCTCCCCTTGAGGAGGAAGAAGAGA-------ATAAAATCTTCACAAGGCTT 1139

Qy 386 eValGluLysLeuLysSer------AsnAspGluLeuLysSerAl 399
              |||||| ||:::::
Db 1140 AAAGGAAAAATCGAAGTCTCCTTCCCCTGTCCCTGACTGAAGATGAAAAGAAGGGCTC 1199

Qy 399 aLeuAspArgIleLysTyrArgValAspAspArgValGluLysTyrLeuAlaProLysProMe 419
              :::   :::||||||||
Db 1200 ACTTGTAGCGCTGCCAGAGCAAACTGCCAGCGAGGAG-------ACTCCTCCACCTTT 1253

Qy 419 t------GluPheAsnProLysProGlyProGlnProGlyTyrPheAlaProArgLysIl 435
              |:::|||
Db 1254 ACTACAACAAGGAGGAGCATCTTCTCCACCACCTCTATCGCCAGCTCCATAGCGAAGAAGAAAT 1313

Qy 435 e------ProThrArgPro------ 439
              |||
Db 1314 AGAGCCCATGGAAGGCCCAGCCCCTGTCCCTGTTCATTCAGTTATCTCCTTCCTCCTAATACAGA 1373

Qy 440 --------ArgLysMetLeu------ 443
              :::|||
Db 1374 TGCTGACACCAGGGAGCTATTAGTATCTCAGCATATCTGTCCAGTTGGTAGGAGGCCTGTC 1433

Qy 444 -ProLeuLeuIleGlySerAsp-ProLys------ValGlnGluGluArgArgH 460
              ||||||| :::||||||
Db 1434 TCCTTTGTCAGTTCCTTCAGACACCAAGACGAATTCCAGCAGAGAAAGTGCCAGAGGA 1493

Qy 460 isProSerThrGluTrp------LysIleAlaLysGluSerArgValLeuT 475
Db 1494 GAGTGTCCTGCCCTGGTTCAGAAAAGCACACTGCTGACTACCAGCCAGCGAGGATCT 1553

Qy   ------P 487
              :::
Db 475 hrAsnLeuLysAsn.------AsnProSerLeuAlaAlaLeu------ 487
              ||||||:::||||||||||
Db 1554 TGAACCTGAGTCAGACAGATCTGCTCAGCCCTCCCTTAAAAAATTGAGGAATTAGCACT 1613

Qy 487 heMetAspAspLysLeuGluAsnThrLeuLysGlyArgGlnMetLeuThrTrpAspGluGlnLeunL 507
              ::: :::|||||||:::
Db 1614 GGGCCAAAGGAATCACTGAACGAATGT-CTGAA-------CAGCCATCTTTCGAACAGAAGG 1666

Qy 507 ysGlyArgThrArgValLysThrIleArgThrAlaLeuProArgLeuPheGlyAlaProThrA 527
              |:||:||| ||| ||||||
Db 1667 AAGGCAGAAGAGCTTCTCATACCCTTCTCCAGCCACAGATTGAAACAGTCAGTCAGTGATT 1726

Qy 527 laLysAlaGlumetIleAspAlaLysAlaLysValPheGlnAspIleGluGluArgProIleProP 547
              :: :::   :::
Db 1727 CATCCTCGACGCCGGTTCCCTCCCTCATCTTCCTCCAGTTCTAGATCAAGATCTCGCTCTC 1786

Qy 547 roLeuPhePheGluProLysGly------ArgHisThrArgLeuArg------ 560
              :::
Db 1787 CT----------GACAGTTCAGGTTCTCGGTCATCACCGTCATCCTCAGATCCAGACAGAGAG 1837

Qy 561 ------TrpThrGlyAlaAsnGluLysGluIleProGlyLeuGlySerArgP 576
              || :::||||||
Db 1838 ATGTAGCCCAGGCACGTACTCACCCTGTGGTAGACCCAAGATGGGCTCCAGAT 1897

Qy 576 heIleLeuProSerLeuAspProThrMetProAlaLeuAsnThrAlaPheSer------ 593
              ::|||
Db 1898 CAACATCAGAGTCCAGATCAAGGTCACGTTCACGTTCTCGTTCAGCATCAAGCAACAGCA 1957

Qy 594 ------ThrGlnGlyValAlaArgAspGluTrpAspPheMetLysIleProA 610
              |||
Db 1958 GAAAATCTCTGAGCCCTGGAGTCTCCAGGGACAGC---AGCACCAGCTATACTGAAACCA 2014

Qy 610 snAsnTrpAsnProGlyAspGluVal 618
              :: |||||||
Db 2015 AAGATCCCTCTTCTTGGTCAGGAGGTT 2040

RESULT 7
AK030316
LOCUS   AK030316   2993 bp   mRNA   linear   HTC 03-APR-2004
DEFINITION   Mus musculus 11 days pregnant adult female ovary and uterus cDNA,
             RIKEN full-length enriched library, clone:5031428K02
             product:similar to CDNA FLJ14503 FIS, CLONE NT2RM1000252, WEAKLY
             SIMILAR TO H.SAPIENS E-MAP-115 MRNA [Homo sapiens], full insert
             sequence.
ACCESSION   AK030316
VERSION   AK030316.1 GI:26326304
KEYWORDS   HTC; CAP trapper.
SOURCE   Mus musculus (house mouse)
  ORGANISM   Mus musculus
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
  AUTHORS   Carninci,P. and Hayashizaki,Y.
  TITLE   High-efficiency full-length cDNA cloning
  JOURNAL   Meth. Enzymol. 303, 19-44 (1999)
  MEDLINE   99279253
   PUBMED   10349636
REFERENCE   2
  AUTHORS   Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
             Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
  TITLE   Normalization and subtraction of cap-trapper-selected cDNAs to
             prepare full-length cDNA libraries for rapid discovery of new genes
  JOURNAL   Genome Res. 10 (10), 1617-1630 (2000)
  MEDLINE   20499374
   PUBMED   11042159
REFERENCE   3
  AUTHORS   Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
             Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M.,
             Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
             Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,

```
                GTTDAGEAAKILAEKRRQARLQKEQEQERLEKEERERLEKEELKRKAEEERLRIEMA
                YKREQEKKRQEEEEKRKAEEKAKEAEELLSKEKQEKKQEQEKKKEAMIEKQKEAA
                EAKAQDAAKQMRLEREQIMLQIEEQERLERKKRIDEIMKTRKSDASLEVKKEDPKVEL
                QPPPDVENKANKAKPVPVPNKIEINVLNTCQKVSGSERAAPETFPQDIFSTGLKPVGGP
                VHLDVLDGKSNSLDDSTEDVQSMDVSPVSKEELISIPEFSPVSEMIPGMSLDQNGTGN
                ARALQDILDFTGPPAFPKKSSENLSLDDCNKNLIEGFNSPGQETTLNTFC"
polyA_signal    2972..2977
                /note="putative"
polyA_site      2993
                /note="putative"
ORIGIN

Alignment Scores:
Pred. No.:                                   Length:               2993
Score:                       0.000839        Matches:              125
Percent Similarity:          155.50          Conservative:         99
Best Local Similarity:       34.89%          Mismatches:           249
Query Match:                 19.47%          Indels:               169
DB:                          4.73%           Gaps:                 25
                             3

US-10-736-868-2 (1-643) x AK030316 (1-2993)

Qy     18 LeuSerAlaGlnSerProSerThrSerAspAlaProGlyAlaLeuLeuSerSerLeuVal  37
             |||   :::  :::   |||   ||| |:||   ::|:||
Db    333 CTGTCTTCATCCACCGTGGCAATATCTTATTCCCCAGACCACCGAGCTCTCGGGAGCCCTCTC 392

Qy     38 GlyLysSerHisGlnLysLeuProLeuAlaProSerMetGluAlaLeuLeuMetGly  57
             ||: ||| ::| ||||
Db    393 AAATCTTCATACAAGTCCTCTCCAACCCGACCACTGAGAAAAAGAAGAATACACCT--  449

Qy     58 ValGlnPheValAspAlaLeuIleLysLysGlyGlnMetGluMetAlaLysGlyAlaPhe  77
             :::|||||     |||   |||
Db    450 ------ATATCTGCCATGGGAGATGCTGGGAAGGGAGCCATGGCTGGACGA------  494

Qy     78 LysThrGlnLeuGluValLeuGluLysValHisProAspGlnPheAspLysTyrLysLys  97
             |||   ||| :::||| :::|::|
Db    495 -------GAGCCTTCACAGATGATGAGAAAATGAGAAAG  524

Qy     98 LeuLysValAspAspLeuAlaAlaAsp----------AlaValMetGlnGlnAlaGlu  113
             :::|||    |||||||
Db    525 GGGCGAGTGGCAACTTCTGCTCTTCTGGGGTCATGGTCCCCACTGAGAAGATGCCGAG  584

Qy    114 ----------------------------------------------MetAlaLys  116
                                                                 ::|||
Db    585 CCTCCTGGAAGATATTTCCAAGGAGTGTGTCTTCTTCCGTGAAATCCAAGATAACTTCAAAA 644

Qy    117 LeuGlnProLysSerGlyAsnAlaPheIleAspMetLeuAsnGlyAsnGlyIle------  134
             ||:::|||   :::
Db    645 ACATATCCACAGTCGCCAAAGACAGCGAAGCCCACATACTTAGGGTCTTCCTGGAAGTAT 704

Qy    135 ----------ProIleGlySerIleArgGlyLeuGluAspAlaIleAlaIleArgGlnArg 151
                       |||   ||:||
Db    705 TACTTTCCACCCATTGCAAACGAAGAAACACCAAAGAAGAAGGCAGAGAAGAGAGAGA   764

Qy    152 AspMetGluAsnThrAspPro----------------SerGluGln 161
             :::
Db    765 AATAAGGAAAAAGGAAGGTGCCCCATGGCTCGCAGAGACACTGTCCTACCTAGAGAAGAAGC 824

Qy    162 IleAlaLysAlaValMetAspLysPheGlnIleThrGlnIleLeuProGlyLeuValAlaAsn 181
             ::: |||   |::|:||||:||
Db    825 CTAGAGAGCGCATGGCTGACAAGTATGCCACCGAG------------------AAG 863

Qy    182 MetIleAlaGlyLysMetProPheLysMetProGlnGlnMetArgLysAlaGlnAlaAla 201
             |||   ::: ::|
Db    864 TATGTAGCAGATAAGCAT--------------------GCAACTGAGAAACATTCTGCC 902

Qy    202 ProSerValPheGlnAlaLeuAlaGlnArgAlaMetLeuGlyLysAsnAlaPro  221
             |||
Db    903 CCTGGACGA----------------------AAGGCTGAGCACAGTGCAGGAAAACCC 938

Qy    222 ValAlaGly-----GlyArgGlyGluGlnLeuArgMetMetAsnArgValAspGln 239
             |||||
Db    939 ACAGCAGGCACCACCGATGCAGGAGGCTGCAAAGATCCTAGCT----------GAA 986
```

                                                                       37
                                                                       392
                                                                       57
                                                                       449
                                                                       77
                                                                       494
                                                                       97
                                                                       524
                                                                       113
                                                                       584
                                                                       116
                                                                       644
                                                                       134
                                                                       704
                                                                       151
                                                                       764
                                                                       161
                                                                       824
                                                                       181
                                                                       863
                                                                       201
                                                                       902
                                                                       221
                                                                       938
                                                                       239
                                                                       986

        Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,
        Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J.,
        Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
TITLE   RIKEN integrated sequence analysis (RISA) system--384-format
        sequencing pipeline with 384 multicapillary sequencer
JOURNAL Genome Res. 10 (11), 1757-1771 (2000)
MEDLINE 20530913
PUBMED 11076861
REFERENCE 4
AUTHORS The RIKEN Genome Exploration Research Group Phase II Team and the
        FANTOM Consortium.
TITLE   Functional annotation of a full-length mouse cDNA collection
JOURNAL Nature 409, 685-690 (2001)
REFERENCE 5
AUTHORS The FANTOM Consortium and the RIKEN Genome Exploration Research
        Group Phase I & II Team.
TITLE   Analysis of the mouse transcriptome based on functional annotation
        of 60,770 full-length cDNAs
JOURNAL Nature 420, 563-573 (2002)
REFERENCE 6 (bases 1 to 2993)
AUTHORS Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Bono,H., Carninci,P.,
        Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hashizume,W.,
        Hayashida,K., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T.,
        Hori,F., Imotani,K., Ishii,Y., Itoh,M., Iwama,A., Kagawa,I., Kasukawa,T.,
        Katoh,H., Kawai,J., Kojima,Y., Kondo,S., Konno,H., Kouda,M.,
        Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M.,
        Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Ohsato,N.,
        Okazaki,Y., Saito,R., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N.,
        Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T.,
        Sogabe,Y., Tagami,M., Tagawa,A., Takahashi,F., Takaku-Akahira,S.,
        Takeda,Y., Tanaka,T., Tomaru,A., Toya,T., Yasunishi,A.,
        Muramatsu,M. and Hayashizaki,Y.
TITLE   Direct Submission
JOURNAL Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of
        Physical and Chemical Research (RIKEN) Laboratory for Genome
        Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
        RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
        Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.jp,
        URL:http://genome.gsc.riken.jp/, Tel:81-45-503-9222,
        Fax:81-45-503-9216)
COMMENT cDNA library was prepared and sequenced in Mouse Genome
        Encyclopedia Project of Genome Exploration Research Group in Riken
        Genomic Sciences Center and Genome Science Laboratory in RIKEN.
        Division of Experimental Animal Research in Riken contributed to
        prepare mouse tissues.
        Please visit our web site for further details.
        URL:http://genome.gsc.riken.jp/.
        URL:http://fantom.gsc.riken.jp/.

FEATURES   Location/Qualifiers
     source   1..2993
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6J"
                     /db_xref="FANTOM_DB:5031428K02"
                     /clone="5031428K02"
                     /sex="female"
                     /tissue_type="ovary and uterus"
                     /clone_lib="RIKEN full-length enriched mouse cDNA library"
                     /dev_stage="11 days pregnant adult"
     CDS   291..1967
                     /note="unnamed protein product; putative
                     similar to cDNA FLJ14503 FIS, CLONE NT2RM1000252, WEAKLY
                     SIMILAR TO H.SAPIENS E-MAP-115 MRNA [Homo sapiens]
                     (SPTR|Q96T17, evidence: FASTY, 73.7%ID, 64.4%length,
                     match=1437)"
                     /codon_start=1
                     /protein_id="BAC26896.1"
                     /db_xref="GI:26326305"
                     /translation="MNLPKQTESPMSKHLSSSTVAISYSPDRALGSPLKSSYKSSPTR
                     TTEKKKNTPISAMGDAGKGAMAGGEPSQMEKMKKGRVATSAASGGHGSPLRRCEPPED
                     ISKRLSSPVKSKITSKTYPQSPKTAKPTYLGSPVKYYFFPIANEETPKKKAEKKRNK
                     EKEGAPGQQSTVLPEESLEKRMADKYATEKYVADKHATEKHSAPGGKAEHSAGKPTA
```

```
Qy   569 IleProGlyLeuGlySerArgPheIleLeuSerLeuAspProThrMetProAlaLeu 588
             |||  |||                                            |||
Db  1914 ATCGAAGGATTTAACAGT----------------------CCCGGCCAAGAAACTACC----CTG 1952

Qy   589 AsnThr 590
             |||
Db  1953 AATACC 1958
```

RESULT 8
BC050866

```
LOCUS       BC050866                7139 bp    mRNA    linear   HTC 25-MAR-2004
DEFINITION  Mus musculus TRAF2 and NCK interacting kinase, mRNA (cDNA clone
            IMAGE:6333274), containing frame-shift errors.
ACCESSION   BC050866
VERSION     BC050866.1  GI:30047190
KEYWORDS    HTC.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 7139)
  AUTHORS   Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,
            Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,
            Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
            Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,
            Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
            Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,
            Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S.,
            Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,
            Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J.,
            McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
            Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,
            Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
            Fahey,J., Helton,E., Ketteman,M., Madan,A., Rodrigues,S.,
            Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,
            Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,
            Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,
            Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smailus,D.E.,
            Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
  TITLE     Generation and initial analysis of more than 15,000 full-length
            human and mouse cDNA sequences
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
   PUBMED   12477932
REFERENCE   2  (bases 1 to 7139)
  AUTHORS   Strausberg,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-APR-2003) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
REMARK      NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT     Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Mark Maconochie, Ph.D. and Nancy L. Freeman,
            Ph.D.
            cDNA Library Preparation: ResGen, Invitrogen Corp
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Genome Sequence Centre,
            BC Cancer Agency, Vancouver, BC, Canada
            info@bcgsc.bc.ca
            Steve Jones, Sarah Barber, Mabel Brown-John, Yaron Butterfield,
            Andy Chan, Steve S. Chand, William Chow, Alison Cloutier, Ruth
            Featherstone, Malachi Griffith, Obi Griffith, Ran Guin, Nancy Liao,
            Kim MacDonald, Amara Masson, Mike R. Mayo, Josh Moran, Ryan Morin,
            Teika Olson, Diana Palmquist, Anca Petrescu, Anna Liisa Prabhu,
            Parvaneh Saeedi, JR Santos, Angelique Schnerch, Ursula Skalska,
            Duane Smailus, Jeff Stott, Miranda Tsai, George Yang, Jacquie
            Schein, Asim Siddiqui, Rob Holt, Marco Marra.
            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAK Plate: 109 Row: c Column: 16
            This clone was selected for full length sequencing because it
```

```
Qy   240 ArgMetGlnArgGluLeuGlnGluGluAspGluAspAspAspLeuGluAspGlu 259
             :::  |||          |||  |||  |||                |||  :::
Db   987 AAGAGACGTCAGGCTCGGCTACAGAGGAACAGAGGAGGCAAGAGCCGACTGGAAAAGGAA 1046

Qy   260 AspValProArgArgSerSerAspGlyProGlnSerGluIleAlaGluHisGlnArg 279
             :::                           |||  :::::::|||     |||
Db  1047 GAACGAGAAAGGCTGGAAAAAAGAGAGAATTGAAAAGAAGGCAGAAGAGGAAAGGCTGCGC 1106

Qy   280 ArgAspLeuAlaArgArgLeuLysSerSerProArgLeuLeuGluLeuGlnLnAsnAla 299
             :::::::|||                :::  :::|||                :::|||
Db  1107 ATAGAATGGCCTATAAACAGAACAGGAAAAAAGGCGCAGGAGGAAGAAGAGAAGAAAGA 1166

Qy   300 GluValGlnSerLeuLeuSerTyrGlnArgMetArgAspSerProLeuSerLysArgArg 319
             :::  :::  |||  :::::::|||                :::
Db  1167 AAGGCAGAAGAG--AAGGCAAAAGAAGAGAAGAGGCTGAGGAGGAGCTCTTGTCAAAAGAGAAAG 1223

Qy   320 ProLeuAlaMetAsnAspGluAspGluSerArgAlaPheArgAlaMet-------------- 334
             |||  |||                    :::|||
Db  1224 CAAGAAAAGGAAAAACACAAGAACAACAAGAAAAAGCCATGATTGAATTGAAAAAACAGAAA 1283

Qy   335 --------GluAlaArgAlaLysLeuAspGlnLysSerGlnLeuValLeuGlyLeuHis 351
                     |||||||::::|||:::  |||
Db  1284 GAAGCAGCAGAAGGCCAG---GATGCTGCTAAACAGATGCGTTA------------- 1331

Qy   352 GlyPheGlyGluSerAspAspArgAspGluAspGluAsnLeuIleAspProSer 371
             |||  |||                   |||  :::
Db  1332 --------------------------------GAAAGAGAACAGATCATG--------- 1349

Qy   372 GluAsnSerPheArgArgAlaProLeuArgLeuSerSerGlyPheValGluLysLeuLys 391
             |||                       :::||||:::
Db  1350 --------CTGCAGATTGAGCAGGAACGACTGGAGGAGGAAGAAG 1385

Qy   392 SerAsnAspGluLeuLysSerAlaLeuAspArgIleLysTyrArgValAspAspValGlu 411
             ||||||       :::     :::
Db  1386 AGATAGATGAAATC---------ATGAAGACAAGGAAGAGTGATGCGTCTTCTAGAA 1436

Qy   412 LysTyrLeuProLysProMetGluPheAsnProLysProGlnProGlyTyrPheAla 431
             :::|||                 |||||
Db  1437 GTGAAGAAGGAAGACCCCAAAGTGGAGCTGCAGCCTCCTCCAGATGTTGAAAATAAAGCA 1496

Qy   432 ProArgLysIleProThrProArgProArgLysMetLeuProLeuLeuIleGlySerAspPro 451
             :::                 :::  |||            :::  :::
Db  1497 AACAAAGCAAAGCCAGTTGTCCCCAACAAGATCGAAATCAATGTGTTGAACACCTGCCAG 1556

Qy   452 LysValGlnGluGluIleArgArgHisProSerThrGluTrpLysIleAlaLysGluSer 471
             |||||                        :::  |||
Db  1557 AAGTAAGTGGCTCAGAGCGTGCTGCTCCAGAACC---------TTTCCCAAGAC 1604

Qy   472 ArgValLeuThrAsnLeuLysAsnAsnProSerLeuAlaAlaLeuPheMetAspAspLys 491
             :::                        ::::|||
Db  1605 ATTTTCTCTACTGGCCTTAAA------CCAGTTGGGGGACCTGTTCATCTGGAT------ 1652

Qy   492 LeuGluAsnThrLeuLysGlyArgGlnMetLeuThrAspGluGlnLnLysGlyArgThrArg 511
             :::::::|||||||:::
Db  1653 --------GTCCTTGATGGGAAATCAAACAGTCTGATGACTCAACTGAACGGTTCAG 1703

Qy   512 ValLysThrIleArgAlaLeuProArgLeuPheGlyAlaProThrAlaLysGluMet 531
             :::                      :::||||:::  :::
Db  1704 TCTATGGATGTG----------------AGTCCTGTTTCCAAGAAGAACTT 1739

Qy   532 Ile--------------------------------------------------------- 532
             |||
Db  1740 ATCTCCATCCCAGAGTTTTCACCAGTGAGTGAAATGATTCCTGGGATGTCTCTGGACCAA 1799

Qy   533 --------AspAlaLysValPheGlnIleAspIleGluGluArgProIleProProLeu 548
                     ::::||||:::         |||
Db  1800 AATGGGAACTGGTAATGCCCGGGCACTTCAAGATATCTTAGATTTCACTGGCCCTCCCGCA 1859

Qy   549 PhePheGluProLysProLysGlyArgHisThrArgLeuArgTrpThrGlyAlaAsnGluLysGlyGlu 568
             |||  |||                      |||:::
Db  1860 TTC-------CCCAAGAAATCCAGTGAAAACCTCAGCCTTGATGACTGACTGTAACAAAAACCTC 1913
```

passed the following selection criteria: Hexamer frequency ORF
analysis
This clone has the following problem: frame shifted.

FEATURES   Location/Qualifiers
source   1..7139
              /organism="Mus musculus"
              /mol_type="mRNA"
              /db_xref="taxon:10090"
              /clone="IMAGE:6333274"
              /tissue_type="Embryo, day 9 mouse (C57BL/6 background)
              otocyst8"
              /clone_lib="NIH_MGC_130"
              /lab_host="DH10B"

ORIGIN   /note="Vector: pCMV-SPORT6.1"

Alignment Scores:
Pred. No.:   0.00476   Length: 7139
Score: 153.50   Matches: 131
Percent Similarity: 34.21%   Conservative: 102
Best Local Similarity: 19.24%   Mismatches: 262
Query Match: 4.66%   Indels: 187
DB: Gaps: 28

US-10-736-868-2 (1-643) x BC050866 (1-7139)

```
Qy    31 AlaLeuLeuSerLeuValGlyLysSerHisGlnLysLeuProLeuAlaProSerMet 50
          ::: ||:    ::  | ::   ||||::    ::   ||| ::
Db  1323 TCATTTATCGAGAGCTGCTTGGTAAAGAATCACAGCCAGCCGGCCACGGAG----- 1376

Qy    51 GluAlaAlaLeuGluLeuMetGlyValGlnPheValAlaAspAlaLeuIleLysLysGlyGlnMet 70
          ::|||||  ||:   ||||||        ::: :||||||||
Db  1377 ------CAGTTGATGAAGCACCCATTCATACGACCAACCTAATGAGACGAGGCAGGTC 1427

Qy    71 GluMetAlaLysLysGlyAlaPheLysThrGlnLeuValLeuGluValLeuValHisProAsp 90
          :::||:||:
Db  1428 ------------------------CGCATCCAGCTGAAG--------------------GAC 1445

Qy    91 GlnPheAspLysTyrLysLysLeuLysValAlaAspAspLeuAlaAlaAspAlaValMetGln 110
          ||  ||:|:   ||||:  ::  :   :||  ||  :::
Db  1446 CACATTGATCGAACAAGAAGAAGGCGAGGAGAAAAAGATGAGACTGAGTATGAATACAGC 1505

Qy   111 GlnAlaGluMetAlaAlaLysLeuGlnProLysSerGlyAsnAlaPheLeuAspMetLeuAsn 130
          ::|||   :||  ::     ||||||    ||: :
Db  1506 GGAAGTGAGGAAGAGGAAGAGGAAGAATGACTCTGGGGAACCC---AGCTCCATTCTGAAC 1562

Qy   131 GlyAsnGlyIleGlySerSerIleArgGlyLeuGlyAspAlaIleArgThrGln 150
          ::||| ||:     ::|::::|:|
Db  1563 ------------CTACCAGGGAGTCAACACTGCGA-----AGGGACTTCCTGAGACTGCAG 1607

Qy   151 ArgAspMetGluAsnThrAspProSerGluGlnIleAlaLysAlaAlaValMetAspLysPhe 170
          ||:|||:  ::  ::
Db  1608 ------CTGGGCCAACAAGGAGGCGCTCAGAGAGCCCCTGCGGGC---------------- 1643

Qy   171 GlnThrGlnIleLeuProGlyLeuValLeuAlaAlaAsnMetIleAlaGlyLysAsnProPheLys 190
          CGCATCCAGGAAGCAGCAG
Db  1643 ------------------------------------------------CGCATCCAGGAAGCAGCAG 1727

Qy   191 MetProGlnGlnMetArgLysAlaGlnGlnAlaAlaProSerValPheGlnGlnAlaLeu 210
          ||||||||: :   ::|||  |||
Db  1644 -----CAACAGCTGGACGAGGCGGGAGAATGAAGAACACAAGCGGCAGCTACTG 1697

Qy   211 AlaGlnArgAlaMetLeuGlyLysAsnAlaAlaProValAlaLeuArgGlyGlyValAlaGlyGlyGluGln 230
          ||:||| ||||::
Db  1698 GCTCGAGCGCCGAGAG-------------------------CGCATCGAAGAGCAG 1727

Qy   231 ArgMetMetAsnArgVal--AspGlnArgMetGlnArgGlnLeuGlnLeuGluGlu 249
          :::    ::|||:    ::||||::||||||||:::
Db  1728 AAGGAGCAAAGCGCAGCGGCTGCAGCGACAACAAAGGCGAGAAAAAGAGCTTCCGGAAACAG 1787

Qy   250 AspGluAspAspAspAspLeuGluAspGluValProArgArgArgSerArgSerAspGly 269
          ||||   :   ::|:
Db  1788 CAGGAGCGGGAGCAGCAGCCGCGCCGGACCTACCAAGACAGATGCGTCGGGAGGAGGAGAGG 1847
```

```
Qy   270 GluProGlnSerGluAlaGluHisGlnAsnArgArgAspLeuAlaArgArgLeuLysSerSer 289
          ::: |||     |||:  ||||:|||  :::
Db  1848 CGTGCCGAACATGAGCAGAGGAATATAAGCCAAACAATTGGAAGAACAGAGACAAGCAGAA 1907

Qy   290 ProArgLeuLysGluLeuGlnAsnAlaGluValGlnLeuSerLeuSerTyrGlnArg 309
          ::|||:   :|||    |||  :::
Db  1908 AGACTGCAGAGGCAGCTAAAGCACAGAGCGGGACTATCTGGTTCCTCCAGCATCAGCGG 1967

Qy   310 MetArgAspSerProLeuSerLysArgArgLeu-----------------AlaMetAsn 324
          ||||| :::     :: ||||||
Db  1968 CAGGAGCAGAGGCCCCTGGAGAG---AAGCCACTGTACCATTACAAGGAGGGCATGAGT 2024

Qy   325 AspGluAspGluSerAlaPhe----ArgAlaMetGluAlaArgAlaLysLeuAspGlnLys 343
          ::: ||||| :::        ::::        ||| :::
Db  2025 CCTAGTGCAGAAGCCGGCCTGGCCAAGGAGGCTCAAGAACGTCAAGACACTCAACCTGACAG 2084

Qy   344 SerGlnLeuValLeuGlyLeuHisGlyPheGlyGlu------SerAspAspAspGluAsp 361
          |||      |||      ||||||||
Db  2085 AGTTCACCTGCCATGCCT---CACAAGGTTGCCAACGGATCTCGGACCCCAACCTGCCC 2141

Qy   362 GluGluAspGluAsnLeu----------------------- 367
          ||||:::
Db  2142 CCAAGATCAGAGTCCTTCAGCATTAGTGGGGTTCAGCCTGCAAGGACACCCCCAATGCTC 2201

Qy   368 ------IleAspPro------------- 370
          ::||||||
Db  2202 AGACCTGTTGACCCCCAGATCCCGCAGCTGGTAGCTGTCAAATCCCAGGGACCTGCCTTG 2261

Qy   371 ---SerGluAsnSerPheArgArgAlaProLeuArgLeuSerGlyPheValGluLys 389
          ::: :||||||  :::
Db  2262 ACGGGCCTCCCAGTCAGTACATGACAACATGAGCAAAGGGGCTGTCTGGTTCCAGGAG--- 2318

Qy   390 LeuLysSerAsnAspGluLeuLeuSerAlaLeuAsnArgIleLysTyrArgValAsp--- 408
          ||||||
Db  2319 -------------GCTCTGAATGTGACCTCTCACCGGTCGAGATG 2351

Qy   409 ---------AspValGluLysTyrLeuAlaProLysProMetGluPheAsnPro 423
          :::      ||||||
Db  2352 CCACGCCAGAACTCGGATCCACCTCCAGAAAACCCTCTCCTCCACAGAATTGAGAAG 2411

Qy   424 LysProGlnProGlyTyrPhe--------------AlaProArgLysIleProThrArg 438
          :::
Db  2412 TTTGACAGAGCTCTTGGTTACGACAGGAAGAAGACATTCACCAAAGGTGCCTCAAAGA 2471

Qy   439 ProArgLysMetLeuProLeuLeu---------------- 446
          ::|||
Db  2472 ACAACTTCTATATCCCCAGCACTAGCCAGAAAGAATTCCCCTGGCAATGGCAGTCGCTCTG 2531

Qy   447 --------IleGlySerAspPro--------LysValGlnGluIleLeuArg---His 460
          :::||||||
Db  2532 GGCCCCAGACTTGGATCTTCAGCCATCAGAGCAAGCAACCCTGATCTGCGCAGGACACAG 2591

Qy   461 ProSerThrGluTrpLysIleAlaGlyLysGluSerArgValLeuThrAsnLeuAsnAsn 480
          :::  ::    :::
Db  2592 CCAGTCCTGGAGAGTTCCCTGCAGCGGACAAGCAGTGGCAGTTCCTCCAGCACT 2651

Qy   481 ProSerLeuAlaAlaLeuPheMetAspAspAspLysLeuGluThrLeuLysGlyArgGln 500
          |||||
Db  2652 CCCAGC-------------TCCCAGCTCCAGTGACTCCCAAGGAGGCTCTCAA 2687

Qy   501 MetLeuThrAspGlnLysGlyArgThrArgValLysThrIleArgAlaLeuProArg 520
          :::
Db  2688 ------CCTGGCTCCCAAGCAGGATCTAGTGAGCGGTCCAACAGTAAG 2741

Qy   521 LeuPheGlyAlaProThrAlaLysAlaGlu------MetIleAspAlaLysValPheGln 538
          ||||||  :::
Db  2742 TCCGAAGGATCACCCGTGCTCCCCATGAGCCTTCCAAGGTGAAACCAGAAGAATCCAGA 2801

Qy   539 AspIle------GluGluArgProLeuProPheGluProLysGlyLysAspArgHis 556
          :::||||||
Db  2802 GACATCACACGGCGCCCCAGTCGGCCACTACGAAGAACAGATGCGTCGGGAGGAGGAGAGG 2861
```

Andy Chan, Steve S. Chand, William Chow, Alison Cloutier, Ruth
Featherstone, Malachi Griffith, Obi Griffith, Ran Guin, Nancy Liao,
Kim MacDonald, Amara Masson, Mike R. Mayo, Josh Moran, Ryan Morin,
Teika Olson, Diana Palmquist, Anca Petrescu, Anna Liisa Prabhu,
Parvaneh Saeedi, JR Santos, Angelique Schnerch, Ursula Skalska,
Duane Smailus, Jeff Stott, Miranda Tsai, George Yang, Jacquie
Schein, Asim Siddiqui, Rob Holt, Marco Marra.

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Series: IRAK Plate: 109 Row: d Column: 16
This clone has the following problem: frame shifted.

```
FEATURES             Location/Qualifiers
     source          1..7139
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:6334359"
                     /tissue_type="Embryo, day 9 mouse (C57BL/6 background)
                     otocysts"
                     /clone_lib="NIH_MGC_130"
                     /lab_host="DH10B"
                     /note="vector: pCMV-SPORT6.1"
ORIGIN
```

```
Alignment Scores:
Pred. No.:              0.00476        Length:       7139
Score:                  153.50         Matches:      131
Percent Similarity:     34.21%         Conservative: 102
Best Local Similarity:  19.24%         Mismatches:   262
Query Match:            4.66%          Indels:       187
DB:                     3              Gaps:         28

US-10-736-868-2 (1-643) x BC050868 (1-7139)

Qy    31 AlaAlaLeuSerSerLeuValGlyLysSerHisGlnLysLeuProLeuAlaProSerMet  50
            ::: ||| :::                  |||||||||||||||  |||
Db  1323 TCATTATCGAGAGCTGCTTGGTAAAGAATCACAGCCAGCGGCCAGCCACCGGAG------ 1376

Qy    51 GluAlaLeuGluLeuMetGlyValGlnPheValAspAlaLeuIleLeuLysGlyGlnMet  70
            :::|||||||   ||||||| ||| ::: |||||||
Db  1377 ------CAGTTGATGAAGCACCATTCATACGAGACCAACCTAATGAGAGGCAGGTC      1427

Qy    71 GluMetAlaLysAlaGlyAlaPheLysThrGlnLeuGluValLeuGlyLysValHisProAsp  90
            ::::||||||          :::
Db  1428 ------------------------------CGCATCCAGCTGAG-------------GAC 1445

Qy    91 GlnPheAspLysTyrLysLysValAlaAspAspLeuAlaAlaAlaAspAlaValMetGln 110
            |||||                   ::: :::
Db  1446 CACATTGATCGAACAAAGAACAGAGAGCGAGGAGAAAAAGATGAGACTGAGTATGAATACAGC 1505

Qy   111 GlnAlaGluMetAlaLysLeuGlnProLysSerGlyAlaAsnAlaPheIleAspMetLeuAsn 130
            ::::|||                   ::: :::
Db  1506 GGAAGTGAGGAAGGAAGAAGAAGAATGAACTGTCGGGGAACCC--AGCTCCATTCTGAAC 1562

Qy   131 GlyAsnGlyIleProIleGlySerSerIleArgGlyLeuGluArgSerIleArgThrGln 150
            |||||
Db  1563 ------------CTACCAGGGGAGTCAACACTGCGA----------AGGGACTTCCTGAGACTGCAG 1607

Qy   151 ArgAspMetGluAsnThrAspProSerGluGlnIleAlaLysAlaAlaValMetAspLysPhe 170
            |||           :::
Db  1608 ------------CTGGCCACACAAGGAGCCGCTCAGAGGCCCTGCCGCGC------------ 1643

Qy   171 GlnThrGlnIleLeuProGlyLeuValAlaLysAsnMetIleAlaGlyLysAsnProPheLys 190
Db  1643

Qy   191 MetProGlnGlnMetArgLysAlaGlnAlaAlaProSerSerValPheGlnAlaAlaLeu 210
            :::             :::
Db  1644 ------CAACAGCTGGACGAGCAGCAGCGCGGGAGAATGAGAACAACAAGCGGCAGCTACTG 1697

Qy   211 AlaGlnArgAlaMetLeuGlyLysAsnAlaProValAlaGlyGlyArgGlyGluGluGln 230
            |||||||||
```

```
Qy   557 ThrArgLeuArgTrpThrGlyAlaAlaAsnGluLysGluIleProGlyLeuGlySerArg--- 575
            ||| :::: |||
Db  2862 CTCCGCATTGAAGAACAACAAACCGCCCCCCTGAAGAAAGTGACTGATTACTTCCTCCAGC 2921

Qy   575 -------- 575
Db  2922 GAGGAGTCGGAGAGCAGTGAGGAAGAGGAAGAGATGAGAGAGTGAGACACATGACGGG 2981

Qy   576 -------PheIleLeuProSerLeuAspProThrMetProAlaLeuAsnThrAla 591
            ::::||   ||| |||||
Db  2982 ACGGTGGCTGCTCAGTGACATACCCAGACTAATACCCACCCGAGCTCCAGGGACAATGAG 3041

Qy   592 PheSerThrGlnGlyArgAlaArgAspGluTrpAspThrMetPheLysIleProAsnAsn 611
            ||| |||          :::
Db  3042 CA-GTACAACATGGGGATGGTCGGGACACATGGCGCTGGAAACTTCGCATGCGGACACCTT 3100

Qy   612 Trp 612
            |||
Db  3101 TGG 3103
```

RESULT 9
BC050868

```
LOCUS       BC050868                7139 bp    mRNA    linear   HTC 25-MAR-2004
DEFINITION  Mus musculus TRAF2 and NCK interacting kinase, mRNA (cDNA clone
            IMAGE:6334359), containing frame-shift errors.
ACCESSION   BC050868
VERSION     BC050868.1  GI:30047833
KEYWORDS    HTC.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
```

```
REFERENCE   1  (bases 1 to 7139)
  AUTHORS   Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,
            Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,
            Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
            Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,
            Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
            Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,
            Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S.,
            Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,
            Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J.,
            McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
            Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,
            Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
            Fahey,J., Helton,E., Ketteman,M., Madan,A., Rodrigues,S.,
            Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,
            Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,
            Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,
            Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smailus,D.E.,
            Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
  TITLE     Generation and initial analysis of more than 15,000 full-length
            human and mouse cDNA sequences
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
   PUBMED   12477932
REFERENCE   2  (bases 1 to 7139)
  AUTHORS   Strausberg,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-APR-2003) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
  REMARK    NIH-MGC Project URL: http://mgc.nci.nih.gov
  COMMENT   Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Mark Maconochie, Ph.D. and Nancy L. Freeman,
            Ph.D.
            cDNA Library Preparation: ResGen, Invitrogen Corp
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Genome Sequence Centre,
            BC Cancer Agency, Vancouver, BC, Canada
            info@bcgsc.bc.ca
            Steve Jones, Sarah Barber, Mabel Brown-John, Yaron Butterfield,
```

```
Db  1698 GCTGAGCGCCAGAAG-----------------CGCATCGAAGAGCAG 1727

Qy   231 ArgMetMetAsnArgVal---AspGlnArgMetGlnIleArgGlnLeuGlnLeuGluGlu 249
             :::         ::|||       ||:|||
Db  1728 AAGGAGCAAAGGCGGAGGCTGGAGGAGCAACAAAGGCGAGAAAAAGAGCTTCGGAAACAG 1787

Qy   250 AspGluAspAspAspAspLeuGluAspGluAspProArgArgSerSerAspGly 269
             |||::                    ||||:|
Db  1788 CAGGAGGCGGGAACAGCGCCGGCACTACGAACAGATGCCGTCGGGAGGAGGAGGAGGAGG 1847

Qy   270 GluProGlnSerGluAlaGluHisGlnAsnArgArgAspArgLeuArgArgLeuLysSerSer 289
             :::  |||||:|||:||     :::   :::
Db  1848 CGTGCCGAACATGAGCAGGAGAATATAAGCGCAAACAATTGGAAGAACAGACAAGCAGAA 1907

Qy   290 ProArgLeuLysGluLeuLeuGlnAsnAlaGluValGlnSerLeuLeuSerTyrGlnArg 309
             :::|||:|       ||
Db  1908 AGACTCGAGAGGCAGCTAAAGCAAGCGGGGACTATCTGGTTCCCTCCAGCATCACGCGG 1967

Qy   310 MetArgAspSerProLeuSerLysArgArgProLeu-------AlaMetAsn 324
             |||||||              :::||||||
Db  1968 CAGGAGCAGCAGACCCCTGGAGAAG----AAGCCACTGTACCATTACAAGGAGGGCATGAGT 2024

Qy   325 AspGluAspGluSerAlaPhe---ArgAlaMetGluAlaArgAlaLysLeuAspGlnLys 343
             :::::  ::  ::|||   |||  |||:::||||||||||:::|||||||
Db  2025 CCTAGTGAGAGCCGCTGGGCCAAGGAGGTAGAAGAACGGTCAAGACTTCAACCGACAG 2084

Qy   344 SerGlnLeuValLeuGlyLeuHisGlyPheGlyGlu------SerAspAspAspGluAsp 361
             |||                              :::      ||:
Db  2085 AGTTCACCTGCCATGCCT---CACACAAGGTTGCCAACAGGATCTCGGACCCCAACCTGCCC 2141

Qy   362 GluGluAspGluAsnLeu 367
             |||:::
Db  2142 CCAAGATCAGATCCTTCAGCATTAGTGGGGTTCAGCCTGCCAGAAGAGACACCCCAATGCTC 2201

Qy   368 ------IleAspPro---------- 370
                   ::::|||
Db  2202 AGACCTGTTGACCCCCAGATCCCGCAGCTCGGTAGCGTCAAATCCCAGGAGACCTGCCTTG 2261

Qy   371 ---SerGluAsnSerPheArgArgAlaProLeuArgLeuSerGlyPheValGluLys 389
                   :::         |||||||       :::
Db  2262 ACGGCCTCCCAGTACATGTACATGAACAACCACCACAAAGGGCCTGTCTCGGGTTCCAGGAG--- 2318

Qy   390 LeuLysSerArgAsnAspGluLeuLysSerAlaLeuAspAspArgIleLysTyrArgValAsp- 408
                                            :::|||::::::::
Db  2319 -GCTCTGAAGTGACCTCTCACCGGGTCGAGATG 2351

Qy   409 ----------AspValGluLysTyrLeuAlaProLysProMetGluPheAsnPro 423
                        |||
Db  2352 CCACGCCGAGAACTCGGATCCCACCTGCAGAAAACCCTCCTCTTCCCCACGAGAGATTGAGAAG 2411

Qy   424 LysProGlnProGlyTyrPhe-------AlaProArgLysIleProThrArg 438
             |||                 |||  :::   ||||||::::::||||||
Db  2412 TTTGACAGAAGCTCTTGGTTACGACAGGAAGAGACATTCCACCAAAGGTGCCTCAAAGA 2471

Qy   439 ProArgLysMetLeuProLeuLeu------------- 446
             |||||||
Db  2472 ACAACTTCTATATCCCCAGACACTAGCCAGAAAGAATTCCCCTGGCAATGGCAGTGCTCTG 2531

Qy   447 -----IleGlySerAspPro------LysValGlnGluIleLeuArg---His 460
                            |||         :::         ||:||||||
Db  2532 GGCCCGCAGACTCGGATCTCAGCCCATCAGACGCAGCACCCTGATCTGCCGCAGGACGAG 2591

Qy   461 ProSerThrGluTrpLysIleAlaLysGluSerArgValLeuThrAsnLeuLysAsnAsn 480
             ||||||   :::         :::
Db  2592 CCCAGTCCTCGAGAGTTCCCTGCAGAGCGGACAAGCAGTGCAGTTCCTCCAGCTCCAGCACT 2651

Qy   481 ProSerLeuAlaLeuPheMetAspAspLysLeuGluAsnThrLeuLysLysGlyArgGln 500
             ||||||                           :::
Db  2652 CCCAGC-----------TCCCAGCCTCCCAGGAGGCTCTCAA 2687

Qy   501 MetLeuThrAspGluGlnLysLysGlyTyrArgThrArgValLysThrIleArgValAlaLeuProArg 520
             :::         :::         ||:||||||
Db  2688 ------CCTGGCTCCCAGCAGGATCCAGTCAGTGAGGATCTCTGCCGGGCCAACAGTAAG 2741
```

```
Qy   521 LeuPheGlyAlaProThrAlaLysAlaGlu------MetIleAspAlaLysValPheGln 538
             |||   :::  ||:|||
Db  2742 TCCGAAGGATCACCCGTGCTCCCCCATGAGCCTTCCAAGGTGAAACCAGAAGAATCCAGA 2801

Qy   539 AspIle------GluGlnArgProIleProLeuPhePheGluProLysGlyArgHis 556
             :::
Db  2802 GACATCACACGGCCGCCCCAGTCGGCCAGCTGATCTGACCGGCATTAGCCAAAGATTACGAGAA 2861

Qy   557 ThrArgLeuArgTrpThrThrGlyAlaAsnGluLysGluIleProGlyLeuGlySerArg-- 575
             |||                        |||:::::
Db  2862 CTCCGGCATTGAAGAAACAAACCGCCCCTGAAGAAAGTGACTGATTACTCTTCCTCCAGC 2921

Qy   575 --------------------------- 575
Db  2922 GAGGAGTCGGAGACAGTAGAGAGAGAGAAGATGGAAGAGTGAGACACATGACGGGG 2981

Qy   576 ------PheIleLeuProSerLeuAspThrThrMetProAlaLeuAsnThrAla 591
Db  2982 ACGGTGGCTGTCAGTGACATACCCAGACTAATACCCAGCGAGCTCCAGGGAACAATGAG 3041

Qy   592 PheSerThrGlnGlyArgArgAlaArgAspGluTrpMetPheLysIleProAsnAsn 611
             |||                  ||||||:::
Db  3042 CA-GTACAACATGGGGATGGTTCGGGACACATGGCCTGGAAACTTCGCATGCCGGACACCTT 3100

Qy   612 Trp 612
             |||
Db  3101 TGG 3103
```

RESULT 10
AY421013
```
LOCUS       AY421013                3918 bp    DNA     linear   GSS 17-DEC-2003
DEFINITION  Mus musculus ACINUS gene, VIRTUAL TRANSCRIPT, partial sequence.
ACCESSION   AY421013
VERSION     AY421013.1  GI:39776970
KEYWORDS    GSS.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 3918)
  AUTHORS   Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejariwal,A.,
            Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
            Ferriera,S., Wang,G.; Zheng,X.H., White,T.J., Sninsky,J.J.,
            Adams,M.D. and Cargill,M.
  TITLE     Inferring nonneutral evolution from human-chimp-mouse orthologous
            gene trios
  JOURNAL   Science 302 (5652), 1960-1963 (2003)
   PUBMED   14671302
REFERENCE   2  (bases 1 to 3918)
  AUTHORS   Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejariwal,A.,
            Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
            Ferriera,S., Wang,G.; Zheng,X.H., White,T.J., Sninsky,J.J.,
            Adams,M.D. and Cargill,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,
            Rockville, MD 20850, USA
COMMENT     This sequence as made by sequencing genomic exons and ordering them
            based on alignment.
FEATURES             Location/Qualifiers
     source          1..3918
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
     gene            <1..>3918
                     /gene="ACINUS"
                     /locus_tag="HCM7415"
ORIGIN

Alignment Scores:
Pred. No.:    0.00247           Length:      3918
Score:        152.50            Matches:     132
```

```
Percent Similarity:   36.02%        Conservative:    96
Best Local Similarity: 20.85%       Mismatches:     230
Query Similarity:      4.63%        Indels:         176
DB:                    9            Gaps:            24

US-10-736-868-2 (1-643) x AY421013 (1-3918)

Qy    23 ProSerThrSerAspAlaProGlyAlaLeuLeuSer---SerLeuValGlyLysSerHis   41
         |||   ::: :::||                     :::  ||  |||||| ||
Db   157 CCGAGTACCTCACGCAGATGGCGGAGCTGGAGGAGGTGACTCTTCGGACGGGAAGCCTCCT  216

Qy    42 GlnLysLeuProLeuAlaProSerMetGluAlaLeuGluLeuMetGlyValGlnPheVal   61
         ||| :::         |||||||  |    ||||||||
Db   217 CAGGGCTCTGCGGGTGACCGACCTGAAGGCCGCACTGGAGCAGCCGAGGC---------  264

Qy    62 AspAlaLeuIleLysLysGlyGln-----------MetGluMetAlaLysGly-  75
         |||         :::              |||||| :::||||
Db   265 -----CTAGCCAAGGCGGCGCAGAAGAGTGCCCTGGTCAAGGCGGCTCAAGGGGGCTCTA  318

Qy    76 -------------------------------AlaPheLysThrGlnLeuGlu  82
                                        |||||| :::
Db   319 ATGCTAGAAAAATTTACAAAAACACTCAACACCCCATGCTGCATTCCAGCCAAATTCACAG  378

Qy    83 ValLeuGluLysValHisProAspGlnPheAspLysTyrLysLysLeuValAlaAspAsp  102
         ::: |||:::::: :::  ||:::   :::  :::|||
Db   379 ATCGGGACGGAGATGAGCGAGCAGACAGCTTCATAAAGCTACCTGGAAAAGCAGCAGGAG  438

Qy   103 LeuAlaAlaAspAlaValMetGlnGlnAla---GluMetAlaLeuGlnLeuProLysSer  121
         |||   ||:   :::  |||           |||  ||| :::::: |||
Db   439 CTGCTGAGGCAGCCCTGAACGGCAGGTCGGGAGCTCGGCGAGCTGGAAGAGCCTCA  498

Qy   122 GlyAsnAlaPheIleAspLeuAsnGlyAsnGlyIle-------------------Pro  135
         |||   ::         :::  ||:::
Db   499 GCTGAGTCGGAGACGAGATGACCCATCCTGAGGGAGTGGCTTCCCTGCCTCCTGAT  558

Qy   136 IleGlySerIleArgGlyLeuGluAspAlaIleArgThrGlnAlaArgAspMetGluAsn  155
         ||||||:::      :::           ||
Db   559 TTTCAGAGCGTGAACAGACAGAGCTGGAGCTCAGCACACATTCACCCGAAAGAGC  618

Qy   156 ThrAspProSerGluGlnIleAlaLeuAlaAlaValMetAspLysPheGlnThrGlnIleIleLeu  175
         ::: |||||||:::
Db   619 TCCTCTTTTTCTGAAGA  636

Qy   176 ProGlyLeuValAlaAsnMetIleAlaGlyTyrLysAsnProPheLysMetProGlnGlnMet  195
         |||:::::: ::           |||||||
Db   637 ---------------------AAAGGAGAATCTGATGATGGAGAAGCCG--  663

Qy   196 ArgLysAlaGlnAlaAlaProSerValPheGlnIleAspLeuAlaLeuAlaLeuAlaArgAlaMet  215
         |||||:::       :::       ::||||||
Db   664 AGGAAAGGGGAAAGACGATCATCCAGAGTT--AGACAGCAAAATCTAAACTCCCTGAG  720

Qy   216 LeuGlyLysLysAsnAlaProValAlaLeuGlyValGlyGluGlnArgMetMetAsn  235
         |||  :::       :::
Db   721 TACAGTCAGACTGCTGAGGAGGAAGAGGATCAAGAGACAGCACCATCTAGAAACCTGAGGGTC  780

Qy   236 ArgValAspGlnArgMet-----------GlnGlnArgGluGlnLeuGluGluGluGluGluAsp  252
         |||
Db   781 AGAGCAGCAGATCGAAATTTAAAAATAGAGGAGGAGAAGAGGAGGAGAAGAGAAGAGAGAT  840

Qy   253 AspAspAspLeuGluGluGluGluProArgArgArgSerAspGlyGluProGln  272
         |||:::||| :::
Db   841 GATGATGAAGGAGGAAGAGGAGGTA-  867

Qy   273 SerGluAlaGluHisGlnIleArgArgAspAspLeueuAlaArgLeuLysSerProArgLeu  292
         |||||:::
Db   868 GATGAGGCCACAGAAATCTAGA---  888

Qy   293 LysGluGluLeuGlnAsnAlaGluValGlnSerLeuLeuSerTyrGlnArgMetArgAsp  312
         |||||:::  :::::::
Db   889 ---------GAGGCTGAGGCCACCAACCCTGAAGCAGTTTGAGGATGAAGAAGGGG  933

Qy   313 SerProLeuSerLysArgArgProLeuLeuMetAsnAspAspGluSerAlaPheArg  332
         ::::::||       ||||||
```

```
Db   934 GAAGAGAGAGACCAGAGCAAACCAGAAAAGTGTGGATGAGAACCCTAAACATCAGA  993

Qy   333 AlaMetGluAlaArgAlaLysAlaLeuAsp-GlnLysSerGlnLeuLeuValLeuGlyLeuLeuHisGl  352
         :::                              :::::::   :::::||:::
Db   994 TCCCAGGA---AAGGGTGAGTTAGAGAAAGAAGGAGGGCGAGTTACTA---------- 1036

Qy   352 yPheGlyValSerAspAspAspGluAspGluAspGlnGluAsnLeuIleAspProSerGl  372
         |||                        :::::: :::||| :::|||::::::::
Db  1037 -------------GATCCCAAGAACAGCTAGAGAAGAAGTCATCTGGCCAGACAG  1077

Qy   372 uAsnSerPheArgArgAlaProLeuArgLeuSerSerGlyPheValGluLysLeuLysSe  392
         |||   :::      |||  |||||::::::   |||
Db  1078 CAACAAGAGAAAGAGACACAAATTGTTCTCC-----------CTCAAGAA  1119

Qy   392 rAsnAspGlyLeuLysSerAlaLeuAspAspArgIleLysTyrArgValAspAspValGluLuY  412
         :::   ::::::            |||::::::::
Db  1120 -GAAAATGAGGTAAAGTCTTCA----------CAAGGTTTAGAGAAAAAATCCCA  1163

Qy   412 sTyrLeuAlaProLysProMetGluPheAsnProLysProGlnProGlyTyrPheAlaPr  432
         :::      :::||| :::::: |||
Db  1164 GTCTCCCTCCCCACCCTGCCTTCCTGAAGACTCTAGAGAAGAGGCCCCTGTTGTGTCAGCC  1223

Qy   432 oArgLysIleProThrArgProArgLysMetLeuProLeuLeuLeu------GlySerAs  450
         |||       :::          |||
Db  1224 AGAACAGATAGTCAGTGAAGAGAAACACCCCACCTTTACTTACCAAGGAAGCGTCATC  1283

Qy   450 pPro-----------------LysValGlnGluIleArg  458
         |||                  :::::::||||||||
Db  1284 TCCACCAACTCATATACAACTCCAGGAAGAGATGGAGGCCAGTGAAGGAGCCACCCCC  1343

Qy   459 ---------------------------------ArgHisProSerThrGl  464
                                          ||||||
Db  1344 TGTCTCATTCAGTTGTCTCCTCCTAACACAGATCGTCGAGCCCACTAGCGTC  1403

Qy   464 uTrpLysIleLysGluSerArgValLeuThrAsnLeuLysAsnAsnProSerLeuAl  484
         |||      :::      |||:::   |||:::
Db  1404 TCCTCACCCTGCGCAGCTCGTGTAAGGAGCCTGTCTCCTTTGTCAGGTACT--------- 1452

Qy   484 aAlaLeuPheMetAspArgAspLysGluGlnThrLeuLysAsnThrLeuLysGlyArgGlnMetLeuThrArLeuThrArAs  504
         |||                  :::||||||
Db  1453 ---------------ACAGACACCAAAGCAGAGTCT----------------------- 1473

Qy   504 pGluGlnLysGlyArgThrArgValLysThrIleArgLysSerThrIleArgArgLeuArgPheGlyAl  524
         |||                 :::::::
Db  1474 ----CCAGCAGGAGAGTGTCAGACGAGAGTGTCCTGCCTCTGGCTCAGAAAAGCTCACT  1529

Qy   524 aProThrArgLysAlaGluMetIleAspAlaLysValPheGlnIleAspIleGluGluArgPr  544
         |||         :::      :::||||||
Db  1530 GCCTGAATGTCTCCCCCAG---------GAGGGTGTTGAAAGTGAGCGTGAAAAAGTC  1577

Qy   544 o-IleProProLeu-----------------PhePheGluProLysGlyArgHisThrArgL  559
         |||                           :::::: |||
Db  1578 TGTCTCCCCTCCCTCTAACAGTGAGCAGTTGGCACCCGCCAAAGGCATCACTGAGGAACC  1637

Qy   559 eu--------------------ArgTrpThrGlyAlaAsnGluLysGluIleProGlyLeuGlySerA  575
         |||                   :::::: |||  ||| :::
Db  1638 TATGAAGAAACAGTCTTTGGAACGAGCAGAGGAGGCAGAAGAGCCTCCCACGCCTCTCCC  1697

Qy   575 rgPheIleLeuProSerLeuAspProThrMetPro  586
         |||      :::
Db  1698 AGAGTACAGTGGGAAGCAGCAGTCAGCCGACTCATCCT  1732
```

RESULT 11
AY414528
LOCUS   AY414528   2934 bp   DNA   linear   GSS 17-DEC-2003
DEFINITION   Homo sapiens RNF20 gene, VIRTUAL TRANSCRIPT, partial sequence,
                genomic survey sequence.
ACCESSION   AY414528
VERSION   AY414528.1 GI:39770490
KEYWORDS   GSS.
SOURCE   Homo sapiens (human)
ORGANISM   Homo sapiens
                Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS   1 (bases 1 to 2934)
             Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejariwal,A.,
             Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
             Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J.,
             Adams,M.D. and Cargill,M.
TITLE   Inferring nonneutral evolution from human-chimp-mouse orthologous
             gene trios
JOURNAL   Science 302 (5652), 1960-1963 (2003)
PUBMED   14671302
REFERENCE   2 (bases 1 to 2934)
AUTHORS   Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejariwal,A.,
             Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
             Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J.,
             Adams,M.D. and Cargill,M.
TITLE   Direct Submission
JOURNAL   Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,
             Rockville, MD 20850, USA
COMMENT   This sequence was made by sequencing genomic exons and ordering
             them based on alignment.
FEATURES   Location/Qualifiers
source   1..2934
             /organism="Homo sapiens"
             /mol_type="genomic DNA"
             /db_xref="taxon:9606"
gene   <1..>2934
             /gene="RNF20"
             /locus_tag="HCM5254"

ORIGIN

Alignment Scores:
Pred. No.:   0.00181
Score: 152.00
Percent Similarity: 39.19%
Best Local Similarity: 19.96%
Query Match: 4.62%
DB:

Length: 2934
Matches: 108
Conservative: 104
Mismatches: 197
Indels: 133
Gaps: 22

US-10-736-868-2 (1-643) x AY414528 (1-2934)

```
Qy     58 ValGlnPheValAspAlaLeuIleLeuLysLysGlyGlnMetGluMetAla----------- 73
             ||  |||  ::: ||||||    ||| :::||||| :::::
Db   1174 GTCTTGTATAATGAGAGCCTACAGTTGAAAGCACACTTGGATGAGGCTCGGACCCTGCTT 1233

Qy     74 ----------LysGlyAlaPheLysThrGlnLeuGluValLeuGluLysValHisProAsp 90
           ::: |||           :::||||||   |||||||||:::||:::
Db   1234 CATGGCACCAGGAACCCACCAGCCACCAGGTTGAGCTTATTGAGCGA----------GAT 1284

Qy     91 GlnPheAspLysTyrLysLeuLysValAlaAspLeuAlaAlaAspAlaAlaValMetGln 110
           ::: ||||||||||||||:::|||::: ||| :::   :::   ::: :::
Db   1285 GAGGTTAGTCTTCATAAGAAGCTGAGACTGAAGTAATTCAGCTAGAAGATACATTGGCC 1344

Qy    111 GlnAla-----------GluMetAlaLysLeuGlnProLysSerGlyAsnAlaPheIle 126
           |||                :::|||::: |||:::|||:::  ::: |||
Db   1345 CAGGTCCGCAAGGAGTGAAATGCTGAGAGAATTTGAGCAGACCCCTTGCT-------- 1398

Qy    127 AspMetLeuAsnGlyAsnGlyIleProIleGlySerSerIleArgGlyLeuGluGluAspAla 146
           ::: |||      :::||||||:::::: |||:::||:::
Db   1399 ----------GCCAATGAACAAGCAGCCGCCTGGTGTTAGGAGGATGCGCCCACCTCATCAGTAGC 1452

Qy    147 IleArgThr---------GlnArgAspMetGluAsnThrAspProSerGluGlnIleIleAla 163
           ::::::              |||    ::: |||:::::::
Db   1453 CTCCAGAATCACCAATCACCAGTGAAAGGGGAGGTCCTGAGATATAAGCGGAAATTGAGA 1512

Qy    164 LysAlaValMetAspLysPheGlnThrGlnIleLeuProGly------LeuValAlaAsn 181
           ::::|||           :::||||||           |||||: |||:::
Db   1513 GAAGCCCAGCCCAGCTTGCAGCTGAACCGCCTGGTCGTAGTGCCTCCTCGCAGTCC 1572

Qy    182 MetIleAlaGlyLysLeuGlyMetProPheLysMetProGlnMetMetArgLysAlaGlnAlaAla 201
           ::: |||:::|||:::::::: |||:::||||||
Db   1573 CAGTCTAGTACTGAGGACCCGAAGGATGAGCCTGCGGAGCTAAAACCAGATTCTGAGGAC 1632

Qy    202 ProSerValPheGlnLeuGlnAlaLeuAlaLeuAlaLysArgAlaMetLeuGlyLysAlaAsnAlaPro 221
           ::: ||| |||
```
```
Db   1633 TTATCCTCC---------CAGTCCTCCAGCTTCAAAGGCATCTCAGAGGATGCCAATGAA 1683

Qy    222 ValAlaAlaGlyGlyArgGlyGluGlnIleArgMetMetAsnArgValAspGlnArgMet 241
           |||   |||||||||:::||||||:::||||||:::
Db   1684 ATCAAGTCTAAACGGATGAAGAAGGAACGAGAAAGGAGGGAAGGAGAGGAGAGGGAA 1743

Qy    242 GlnGlnArgGlu-------------------------------LeuGlnGlnGlu 248
           :::::||||||                                 ||||||
Db   1744 CGAGAAAGAGAACGGGAGAAGGAGGAGAGAAGAAGACGAGAAGCTAAAAGAG 1803

Qy    249 GluAspGluAspAspAspProLeuGlyLysGluAspGluAspValProArgArgSerSerAsp 268
           :::||||||::: |||  ::: ::: :::  :::
Db   1804 TCAGAAAAAGAGAGAGATTCTGCTAAGGATAAAGAG------AAAGGCAAACATGATGAT 1857

Qy    269 GlyGluProGlnSerGluAlaGlu------------HisGlnArgArgAspLeuAlaAlaArg 285
           |||            ::: |||||||            ::: :::||||||::: :::
Db   1858 GGACGG--AAAAAGGAAGCAGAAATTATCAAACAATTGAAGATTGAACTCAAGAAGCA 1914

Qy    286 LeuLysSerSerProArgLeuLysGluLeuLeu--------- 296
           ::::||  |||||||   :::||
Db   1915 CAGGAGGCCAAAAGGAGCAGAAACTATTGCTGGGATATGTACCGTTCTGCCCAAAGGAA 1974

Qy    297 ------------------------GlnAsnAlaGluValGlnSer 303
                                    :::
Db   1975 CAGGAGAGACAAAGTTCAGCTGCAGCTGAGAAGAAGTCTAAGGCAGAGTTGGAAGAT 2034

Qy    304 LeuLeuSerTyrGlnArgMetArgAsp---SerProLeuSerLysArgArgProLeuAla 322
           |||                :::::: :::||
Db   2035 CTA---AGGCAAAGACTCAAGGATCTCAAGGAAGAAGGAAGAAAGAGAAGAGAACAAGAAA 2088

Qy    323 MetAsnAspGluAspGlu--SerAlaPheArgAlaMetGlu----------- 335
           |||   :::||||||::::::|||
Db   2089 ATGGCTGATGAGGATGCCTTGAGGACATCCGGGCAGTGAGATGAATACCTA 2148

Qy    336 -----------AlaArgAlaLysAspGlnLysSerGlnLeuValLeuGlyLeuLeuHis 351
                      :::||||||:::::||:::
Db   2149 CAGAAGAAGCTAGCCATGGCCAAGCAGCAGAAGAAGCA---CTCCTCCTGAAATGGAT 2205

Qy    352 GlyPheGlyGluSerAspAspAspGluGluGluAsp----------GluAsnLeu 367
           |||   ::: :::|||:::::: :::
Db   2206 GTCACAGGGCCAGGCCTTGAAGACATGCAGGAGCAAAATATCCGTTTGATGCAGCAATTG 2265

Qy    368 IleAspProSerGluAsnSerPheArgArgAlaProLeuArgLeuSerSerGlyPheVal 387
           ::::::|||         :::::: |||::::::
Db   2266 CGGGAAGGATGATGCAAATTTCAAGCTCATGTCCGTACAAGTCCAATCAGATC 2325

Qy    388 GluLysLeu------------LysSerAsnAspGluLeuLysLysSerAlaLeuAspArgIleLysTyr 405
           ::: ::::::              ||:::
Db   2326 CATAAGTTGCTTAAAGAAGAGAAGGAGGAGCTGGCAGACCAGGTGTTGACTCTGAAGACT 2385

Qy    406 ArgValAsp----------- 408
           :::||||||
Db   2386 CAGGTTGATGCCCAGCTACAGGAGGAAGAACTGGAGAAGAAGAGCATCGTTACAG 2445

Qy    409 -----------AspValGluLysTyrLeuAlaProLysProMetGluProAsn 422
                      :::   |||     :::||
Db   2446 AGCAACATTGGCACAGGGACAGAGAAAGACGCTTGGGTTCTTAGGACCCAAGCCTTAGAGATGAAT 2505

Qy    423 ProLysProGlnProGlyTyrPheAlaProArgLysIleProThrArgProLysMet 442
           |||      :::
Db   2506 AAACGC------------------------AAGGCAATG 2520

Qy    443 LeuProLeuLeuIleGlySerAspProLysValGlnGluIleArg-ArgHisPro-- 461
           :::            ::: ||||||:::||||| ||:::
Db   2521 GAGGCAGCCCAGCTTGCAGATGACCTCAAAGCAACTGGAGTTGGCTCAGAAGAAGCTA 2580

Qy    462 ------------SerThrGluTrpLysIleAlaLysGluSerArgValLeuThrAsnLeu 477
                      :::
Db   2581 CATGATTTTCAGGATGAGATCGTGGAACAGTGTTACCAAAGAAAAAGGACACATGTTCAT 2640

Qy    477 uLysAsnAsnProSerLeuAlaAlaLeuPheMetAspAspLysLeuGlnAsnThrLeuLeu 497
           :   |||           ||||||
```

```
Db   2641 TTCAAACGAGCCCAGGAGGACATCTCTAGA-CTTCGCAGGAAGCTGAGACCACAAAGAA 2699
Qy    497 B 497
Db   2700 A 2700

RESULT 12
CL960352
LOCUS       CL960352      3012 bp    DNA    linear    GSS 21-SEP-2004
DEFINITION  OsIFCC004301 Oryza sativa Express Library Oryza sativa (indica
            cultivar-group) genomic, genomic survey sequence.
ACCESSION   CL960352
VERSION     CL960352.1    GI:52375402
KEYWORDS    GSS.
SOURCE      Oryza sativa (indica cultivar-group)
ORGANISM    Oryza sativa (indica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 3012)
AUTHORS     Ma,L., Wang,C., Chen,C., Liu,X., Su,N., Li,L., Wang,X., Cao,M.,
            Jiao,Y., Sun,N., Zhang,X., Bao,J., Sun,D., Zhao,H., Yuan,L.,
            Wong,G.K.S., Deng,X.W. and Wang,J.
TITLE       An analysis of transcriptional regulation of the rice genome and
            its comparison to Arabidopsis
JOURNAL     Unpublished (2004)
COMMENT     Contact: Chen Chen
            Department of Bioinformatic
            Beijing Institute of Genomics
            Chinese Academy of Sciences, Beijing 101300, China
            Tel: 86-10-80481559
            Fax: 86-10-80488676
            Email: chenchen@genomics.org.cn
            Rice genomic sequence.
            Class: exon-trapped.
FEATURES    Location/Qualifiers
     source 1..3012
            /organism="Oryza sativa (indica cultivar-group)"
            /mol_type="genomic DNA"
            /db_xref="taxon:39946"
            /clone_lib="Oryza sativa Express Library"
            /note="Oryza sativa exon trapped genomic sequences "
ORIGIN

Alignment Scores:
Pred. No.:              0.00236
Score:                  151.00
Percent Similarity:     34.98%          Length:           3012
Best Local Similarity:  20.65%          Matches:          121
Query Match:            4.59%           Conservative:     84
DB:                     9               Mismatches:       233
                                        Indels:           148
                                        Gaps:             25

US-10-736-868-2 (1-643) x CL960352 (1-3012)

Qy    24 SerThrSerAspAlaProGlyAlaLeu------------------LeuSerSerLeu 36
Db   589 TCGACGAATGCTGCTATCGCGATGCTGCCTAGAGGAGCAGTCTTGGATTGTCCAAGTTG 648

Qy    37 ValGlyHisSerHisGlnLysSerGlnLeuLeuProLeuAlaProSerMetGluAlaLeuGlu------ 54
Db   649 GTGGATGAGATATGAAAAGCTGAAGGGAATGAAGGGCGTGTCGGGCACTTCGGTCCGG 708

Qy    55 ------------------LeuMetGlyValValGlnPheValAspAlaLeu 64
Db   709 AGTTCTGTGCTGCTCATTGGGCAGAGAGATGATGTATGGCAATCAGTTCCGGATGCTGAT 768

Qy    65 IleLeuLysGlyGlnMetGluMetAla------------------ 73
Db   769 GTCTTGGAAGATGAATCAGAGACGACGGCTCTTGTTGCTGGGAGGTGAGGAGATTTGAAGGTG 828

Qy    74 ------------LysGlyAlaPheLysThrGlnLeuGluValLeuGluValLysValHis 88
Db   829 ATTCCTCTGAGAATGCGGTGACCTCGAGTACGCGTAGAACTGGTAGAAAGAAAATCCAC 888

Qy    89 ProAspGlnPheAspLysTyrLysLeuLysValAspAspLeuAlaAlaAspAlaVal 108
Db   889 GAGAGGATTACTGCTGCTATTTATTCACGTTATCAGTTCTAGAA---GCTCCGGAGCTGAA 945

Qy   109 MetGlnGlnAlaGluMetAlaLysLeuGlnProLysSerGlyAsnAlaPheIleAspMet 128
Db   946 GCTCAAGTAAATGATATGAGGAAGCATCACTGAAGTTAAGCAAAGCA--- 993

Qy   129 LeuAsnGlyAsnGlyIleProIleGlySerSerIleArgGlyLeuGluAspAlaIleArg 148
Db   994 CTGAACCTGGAAGTATTAAA----TCATTGGTGGAAAGAGAGCGACCAAAAGAGCAAC 1047

Qy   149 ThrGlnArgAspMetGluAsnThr------------------AspProSerGluGlnIleAla 163
Db  1048 ATTGAACGGCGTGCCAAAAATACTGGATCAACAGCTAAAGAGCCAATGCAAGAGATGGTG 1107

Qy   164 LysAlaValMetAspLysPheGlnThrGlnIleLeuProGlyLeuValAlaAsnMetIle 183
Db  1108 AAAAGCAATAACGAT---------------------ACTGGGATTATTGAGAATGTGGAT 1146

Qy   184 AlaGly----------LysAsnProPheLysMetProGlnGlnMetArgLysAlaGlnAla 200
Db  1147 GACTCCAGCTGCAAAAAACACCTCAACTAATGAAAAAGATACTCAGAAAGCACAGAAA 1206

Qy   201 AlaProSerValPheGlnGlnAlaLeuAlaGlnAlaArgAlaMetLeuGlyLysAsnAla 220
Db  1207 CAAGTTGAAAAGGAGCTAAAACAGAAGGAGAAAGAAGAAGCTCGAATGAGAAAAACAACA 1266

Qy   221 ProValAlaGlyGlyArgGly----------GluGluGlnArgMetMetAsnArgValAsp 238
Db  1267 AAGAAAACAGAAGAGAGCGCTGAGACAAACAGAAACGTGAAAAGAAGAAGCTGAA 1326

Qy   239 GlnArgMetGlnGlnArgGluLysGlnArgLeuGlnGluGluLysAspAspLeuGluAsp 258
Db  1327 ATGAAGAAACAGCAGGAAGGAGAAGCAGAGAAAGGAGCACAGAAAGAAGAACAGAGAGGCGGCCGTGAA 1386

Qy   259 GluAspValProArgArgSerSerAspGlyProGlnLeuGlnSerGlyLeuGluHisGln 278
Db  1387 AAGGAAGAAGCTGAAACGAGGAGCAACAAAAGAGCAGCAAGAGGAAGCTGAGAAAGAA 1446

Qy   279 ArgArgAspLeuAlaArgLeuLysSerSerProArgLeuLysSerGluLeu------ 296
Db  1447 CAGAAG------CGCGTGAGAAGGAAGCTGTCCAACTCAAGAGCAACTGGCCATT 1497

Qy   297 ------------------GlnAsnAlaGluValGlnSerLeuLeu 305
Db  1498 CAGAAGCAAGCTTCTATGATGGAACGGTTTTCAAAAATAAGAAAAACAGTGAAAAGCTT 1557

Qy   306 SerTyrGlnArgMetAspProLeuSerLysArgArgProLeuAlaMetAsnAspAsp 325
Db  1558 GAGAAATCTGGAGGAAAAGGACTCAGGCGTTCAAACTACTGATCCTGACCACCAACAA 1617

Qy   326 Glu------------------AspGluSerAlaPheAlaArgAlaMetGluAlaArg 337
Db  1618 GAGGTGGTACCATGGTTACATCCATAATTGATTCTTCATTCTTCAAAAAGAAAACTGG 1677

Qy   338 AlaAlaLysLeuAspAsnLysSerGlnLeuSerGlnLeuGluValLeuGlyLeuHisGly 357
Db  1678 GCTTTGGAGGATCTTCGCAGATTCCAGATTCAGATCAGTGGTTGGCAGAAGTTATCTAGCTATAAC 1737

Qy   358 AspAspGluAspAlaGluLysGluAspGluAsnLeuIleGluAspProSerPheAlaLeu 377
Db  1738 AGGTCCTCCAGGTGGGGGATCAGAAC------AAACCCAAGAAGAAAGAGCTTTTAAGGAG 1791

Qy   378 AlaProLeu--ArgLeuSerGlyPheValGluLysLeuLysSer------ 392
Db  1792 CTCAAACTTCAAAAACCTGATAACATGCTTGAGGAAATTCTTCTCCTAATGAAGAT 1851

Qy   393 ------------------AsnAspGluLeuLysSerAlaLeuAspArgIleLys 404
Db  1852 ACCTGCCACAAATTTAAGTCAGGAAAATGAGCCAGATAAGTCAGCAAATGAT------ 1902
```

```
Qy   148  --------------------------------------------------  148
Db   144  GluAspAlaIleArgThrArgAspMetGluAsnThrAspProSerGluGlnIleAla  163
                                                         |||:::||
Db   149  ATGGACAGAGCTGAGCCGGCTGAGTCGAGTCAGACAAG  181

Qy   164  LysAlaValMetAspLysPheGlnThrGlnIleLeuProGlyLeuValAlaAsnMetIle  183
               ::: 
Db   182  AGCCACCGCTTGAGGCTGACCTGTGTG  226

Qy   184  AlaGlyLysAsnProPheLysMetProGlnMetArgLys----------AlaGlnAla  200
               :||::
Db   227  GCA  227 ... CCACAGAGAAGCCGAAGCCACTGAGGCTGAGTTG  265

Qy   201  AlaProSerSer---ValPheGlnAlaLeuAlaLeuArgAlaMetLeuGlyLysAsn  219
          ||||||||||              |||||||
Db   266  GCCCCGGTCCTTGACGCGCCTGACGCGTGCCCCCGGCTGCGGCTCCGTGAGAAGGAG  325

Qy   220  AlaProValAlaGlyValGlyArgGlyGluGlnLeuArgMetMetAsnArgValAspGln  239
          |||              :::     :::
Db   326  GCCCACAGATGCTGAGGGTGATGTCGCTTCCCGAACAGGCGGTTGCTCCCGTGGTTGAGGAG  385

Qy   240  --------ArgMetGlnLeuArg------------------GluLeuGlnGluGluAsp  250
                  :::||||||                      |||:::
Db   386  GAGTTGGACCGCCCACAGGAGCGTCTGCCACAGCTCTGACCAAGTTGGAGGAGGCCAGAGAG  445

Qy   251  GluAspAspAspAspProLeuGlu-----AspGluAspValProArgArgSerSerAsp  268
          :::                            |||
Db   446  AAGGCTCGTGATGAGTGAGAGGAGGAGGAGCAAGGAAGCAGGGCCAGGAAGGAAGCAGGAC  505

Qy   269  GlyGluProGlnSerGlnValAlaGluHisArgArgAspLeuAlaArgLeuAlaArgLeuLysSer  288
          |||     |||:::|||
Db   506  GAGGAGAAGATGAGCGGCGGAAGAACAACCCAAGGAGGAGGAGGCCAAACAACAAGGACAGAGAG  565

Qy   289  SerProArgLeuLysAspLeuLeuLeuGlnAlaAsnAlaGluValGlnSerLeuSerLeuSerTyrGln  308
          :::||                    GAGGGGCCCGAAAGCGGGGGACAAAGAGAGGAGGAA
Db   566  GCTCGACCGCAAAAAGAG--------GAGGGGCCCGAAAGCGGGGGACAAAGAGAGGAGGAA  619

Qy   309  ArgMetArgAspSerProLeuSerLysArgArgProLeuAlaMetAsnAspGluAspGlu  328
          |||                                    :::|||
Db   620  CGGGAACCGAACAGAGGAGGCGGCGGGAGCGGGAGCAAAAGCCCAGCCGGCGGAGGAG  679

Qy   329  SerAlaPhe------------------ArgAlaMetGluAlaArgAlaLysLeuAspGln  342
          |||                       :::||||||
Db   680  GAGGCGGAAAACGGGCAGAACAACCCAAGGAGGCCGGGGAGGAGGCGGAGAAGAACCCA  739

Qy   343  LysSerGlnLeuValLeuGlyLeuHisGlyPheGlyGluSerAspAspAspAspGluGlu  362
               |:::                                  CAAAAGGAAGACAAAAACGAGGAG
Db   740  ---------------CAAAAGGAAGACAAAAACGAGGAG  763

Qy   363  GluAspGluAsnLeuIleAsp---ProSerGluAsnSerPheArgArgAlaProLeuArg  381
          :::                       |||
Db   764  GAAACCAAGGAGCCACCAGACAAGCCCAACAAGCCCGAGACCCGGCGGCCCCAG  823

Qy   382  LeuSerSerGlyPheValGluLysSerLeuLysSerAsnAspGluLysLysSerAlaLeuAsp  401
               |||                                  :::
Db   824  AGGCCCGCCGCCTGAGACGCCGAGAAGAGCAGAACAACCCGAGACGAGAAAGCCGAAAGCCCCAG  883

Qy   402  ArgIleLysTyrArgValAspAspValGluLysTyrLeuAlaProLysProMetGluPhe  421
          :::     |||:::
Db   884  AAACGGAAGAACAAGGCCGAGGAGCGCCCCCAACCGACCAAGAACTCCC  943

Qy   422  AsnProLysProGlnPro----------GlyTyrPheAlaPro----  432
          |||                         |||
Db   944  AACAAAACCCAACCCCCCCAGCACGCGGGGGAGGGCAACAAGGCACCCCCGCG  1003

Qy   433  ---ArgLysIleProThrArg---ProArgLysMetLeuProLeuLeuIleGlySerAsp  450
          |||              |||
Db   1004 GGGCGAGCGCACCCAACAGAGGGCCCGGATCCAACACCCCGCACACAGGG---CCG  1060

Qy   451  ProLysValGlnGluIleLeuArgArgHisProSerThrGluTrpLys  466
          |||:::
```

```
Qy   405  TyrArgValAspAspVal----------  410
               |||:::
Db   1903 GTTGATATGCTACCAGCGGTTGAGCTAGCAGTTTCATGGCACCAATCATGCTAAT  1956

Qy   411  --------GluLysTyrLeuAlaProLeuProMetGluPheAsnProLysProGlnPro  427
               :::||
Db   1957 CCTCTGCCAACTAGATCAATAAAGCGAAGCTTTGCAATTTGACAAAAGCAACAGACCT  2016

Qy   428  GlyTyrPheAlaPro---ArgLysIle--------ProThrArgProArgLys  441
          |||
Db   2017 GCATATTATGGGACTTGGAGGAGAAAAGTGCTGTTCGTTGGCCCAAGATGTCCACTTAAG  2076

Qy   442  MetLeuProLeuLeuIleGlySerAspProLysValGlnGluIleLeuArgArgHisPro  461
          |||                       :::|||
Db   2077 ATG----------------------GATCCTGATCTTGACATATGAGGTT------GATAGT  2109

Qy   462  SerThrGluTrpLysIleAlaLysGluSerArgValLeuThrArgAsnLeu---LysAsnAsn  480
               ||                        :::|||
Db   2110 GATGATGAATGGAAGAAGAGGAGGATCCTGGTCGAGAGCCTTTCCGACTGCGAGAAGGACAAT  2169

Qy   481  ProSerLeuAlaAlaLeuPheMetAspAspProAspLysLeuGluAsnThrLeuLysGlyArgGln  500
          |||                                  :::|||
Db   2170 -----------------GATGAAGTTATGGAGGAGGATTCCAAG--------  2196

Qy   501  MetLeuThrAspGluGln  506
          :::|||||||
Db   2197 ---ATCACAGATGAAGAG  2211
```

RESULT 13
```
LOCUS       CR679256               1128 bp    mRNA      linear      HTC 19-AUG-2004
DEFINITION  Tetraodon nigroviridis full-length cDNA.
ACCESSION   CR679256
VERSION     CR679256.1    GI:51177163
KEYWORDS    HTC; cDNA; full-length; Tetraodon nigroviridis.
SOURCE      Tetraodon nigroviridis
ORGANISM    Tetraodon nigroviridis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
            Tetraodontoidea; Tetraodontidae; Tetraodon.
REFERENCE   1  (bases 1 to 1128)
AUTHORS     Genoscope.
TITLE       Direct Submission
JOURNAL     Submitted (10-AUG-2004) Genoscope - Centre National de Sequencage -
            2 rue Gaston Cremieux, CP 5706 - 91057 EVRY cedex - FRANCE
            (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT     The sequences are based on single pass reads.
            More information available at
            http://www.genoscope.cns.fr/tetraodon.
FEATURES             Location/Qualifiers
     source          1..1128
                     /organism="Tetraodon nigroviridis"
                     /mol_type="mRNA"
                     /db_xref="taxon:99883"
                     /tissue_type="Muscle"
ORIGIN
```

Alignment Scores:
Pred. No.:	0.000623	Length:	1128
Score:	150.50	Matches:	90
Percent Similarity:	36.62%	Conservative:	55
Best Local Similarity:	22.73%	Mismatches:	156
Query Match:	4.57%	Indels:	95
DB:	3	Gaps:	17

US-10-736-868-2 (1-643) x CR679256 (1-1128)

```
Qy   104  AlaAlaAspAlaValMetGlnGlnAlaValMetAlaLeuAlaGluMetAlaLysLeuGlnProLysSerGlyAsn  123
          |||     :::         ||||||
Db   92   GCCATGGATGCCATCAAGAAGATGCAGATGCGAGATGTCTCAAGCTGCCACAAGGAAACTGCC---  148

Qy   124  AlaPheIleAspMetLeuAsnGlyAsnGlyIleProIleGlySerSerIleArgGlyLeu  143
```

```
Db          1061 CCGCGC----------------CGGCACCACCACGAGCACTCACAGCAAG 1093

RESULT 14
AK030449
LOCUS        AK030449                3431 bp    mRNA    linear    HTC 03-APR-2004
DEFINITION   Mus musculus adult male pituitary gland cDNA, RIKEN full-length
             enriched library, clone:5330413K22 product:unknown EST, full insert
             sequence.
ACCESSION    AK030449
VERSION      AK030449.1  GI:26081823
KEYWORDS     HTC; CAP trapper.
SOURCE       Mus musculus (house mouse)
ORGANISM     Mus musculus
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE    1
AUTHORS      Carninci,P. and Hayashizaki,Y.
TITLE        High-efficiency full-length cDNA cloning
JOURNAL      Meth. Enzymol. 303, 19-44 (1999)
MEDLINE      99279253
PUBMED       10349636
REFERENCE    2
AUTHORS      Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
             Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
TITLE        Normalization and subtraction of cap-trapper-selected cDNAs to
             prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL      Genome Res. 10 (10), 1617-1630 (2000)
MEDLINE      20499374
PUBMED       11042159
REFERENCE    3
AUTHORS      Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
             Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M.,
             Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
             Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
             Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,
             Yoneda,Y., Ishikawa,T., Ozawa,A., Tanaka,T., Matsuura,S., Kawai,J.,
             Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
TITLE        RIKEN integrated sequence analysis (RISA) system--384-format
             sequencing pipeline with 384 multicapillary sequencer
JOURNAL      Genome Res. 10 (11), 1757-1771 (2000)
MEDLINE      20530913
PUBMED       11076861
REFERENCE    4
AUTHORS      The RIKEN Genome Exploration Research Group Phase II Team and the
             FANTOM Consortium.
TITLE        Functional annotation of a full-length mouse cDNA collection
JOURNAL      Nature 409, 685-690 (2001)
REFERENCE    5
AUTHORS      The FANTOM Consortium and the RIKEN Genome Exploration Research
             Group Phase I & II Team.
TITLE        Analysis of the mouse transcriptome based on functional annotation
             of 60,770 full-length cDNAs
JOURNAL      Nature 420, 563-573 (2002)
REFERENCE    6   (bases 1 to 3431)
AUTHORS      Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Bono,H., Carninci,P.,
             Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hashizume,W.,
             Hayashida,K., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T.,
             Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kasukawa,T.,
             Katoh,H., Kawai,J., Kojima,Y., Kondo,Y., Konno,H., Kouda,M.,
             Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M.,
             Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Ohsato,N.,
             Okazaki,Y., Saito,R., Saitoh,H., Sakai,C., Sakai,Y., Sakazume,N.,
             Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T.,
             Sogabe,Y., Tagami,M., Tagawa,A., Takahashi,F., Takaku-Akahira,S.,
             Takeda,L., Tanaka,T., Tomaru,A., Toya,T., Yasunishi,A.,
             Muramatsu,M. and Hayashizaki,Y.
TITLE        Direct Submission
JOURNAL      Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of
             Physical and Chemical Research (RIKEN), Laboratory for Genome
             Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
             RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
             Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.jp,
             URL:http://genome.gsc.riken.jp/, Tel:81-45-503-9222,
```

```
COMMENT           Fax:81-45-503-9216)
                  cDNA library was prepared and sequenced in Mouse Genome
                  Encyclopedia Project of Genome Exploration Research Group in Riken
                  Genomic Sciences Center and Genome Science Laboratory in RIKEN.
                  Division of Experimental Animal Research in Riken contributed to
                  prepare mouse tissues.
                  Please visit our web site for further details.
                  URL:http://genome.gsc.riken.jp/
                  URL:http://fantom.gsc.riken.jp/.
FEATURES          Location/Qualifiers
     source       1..3431
                  /organism="Mus musculus"
                  /mol_type="mRNA"
                  /strain="C57BL/6J"
                  /db_xref="FANTOM_DB:5330413K22"
                  /db_xref="taxon:10090"
                  /clone="5330413K22"
                  /sex="male"
                  /tissue_type="pituitary gland"
                  /clone_lib="RIKEN full-length enriched mouse cDNA library"
                  /dev_stage="adult"
     misc_feature 1..3431
                  /note="unknown EST (GB|AI593306, evidence: BLASTN, 98%,
                  match=441)"
ORIGIN

Alignment Scores:
Pred. No.:                  0.0032            Length:            3431
Score:                      150.50            Matches:           134
Percent Similarity:         35.92%            Conservative:      93
Best Local Similarity:      21.20%            Mismatches:        230
Query Match:                4.57%             Indels:            176
DB:                         3                 Gaps:              25

US-10-736-868-2 (1-643) x AK030449 (1-3431)

Qy     24 SerThrSerAspAlaProGlyAlaLeuLeuSer---SerLeuValGlyLysSerHisGln  42
          |||||||              |||               :::|||  |||||
Db      4 AGTACCTCACGCAAGATGGGGAGCTGGAGGAGGTGACTCTGGACGGGAGGCCTCTTCAG  63

Qy     43 LysLeuProLeuAlaProSerMetGluAlaLeuMetGlyValGlnPheValAsp  62
          ||| ::                           ||||||||
Db     64 GCTCTGCGGGTGACCGACCTGAAGGCCGACTGAAGCAGCGAGGC-----------  108

Qy     63 AlaLeuIleLysLysGlyGln-------------MetGluMetAlaLysGly-------  75
          |||||||||| |||                   :::::::
Db    109 ---CTAGCCAAGAGCGGGCAGAAGAGTGCCCTGGTCAAGCGGCTCAAGGGGCTCTAATG  165

Qy     76 --------------------------------AlaPheLysThrGlnLeuGluVal  83
                                          |||||||    :::::::
Db    166 CTAGAAAATTACAAAACACTCAACACCCCATGCTGCATTCCAGCCAAATTCACAGATC  225

Qy     84 LeuGluLysValHisProAspGlnPheAspLysTyrLysLysValAspAspLeu  103
          |||:||: ::                    :::
Db    226 GGGGAGGAAATGAGCCAGAACAGCTTCATAAAGCAGTATCTGGAAAAGCAGCAGGAGCTA  285

Qy    104 AlaAlaAspAlaValMetGlnIleAla---GluMetAlaLysLeuGlnProLysSerGly  122
               :::   ||||||||   :::::::||||:::|||::|||
Db    286 CTTAGGCAGCGCCTGGAACGGGAGCTCGGAGGCGTCGCGGAACTTGAAGAAGCCTCAGCT  345

Qy    123 AsnAlaPheIleAspMetLeuAsnGlyAsnGlyIle----------------ProIle  136
          :::    :::   |||      ::: ||||:
Db    346 GAGTCGGAGGACGGACGATGACCCATCCTGAGGGAGTGGCTTCCCTGCTGCCTCCTGATTTT  405

Qy    137 GlySerSerIleArgGlyLeuGluAspAlaIleArgThrGlnArgAspMetGluAsnThr  156
          ||||:::|||:::
Db    406 CAGAGCAGCCTGAACAGACCCAGAGCTGGACTCAGCACACATTCACCCCGAAAGAGCTCC  465

Qy    157 AspProSerGluGlnIleAlaLysAlaValMetAspLysPheGlnThrGlnIleLeuPro  176
          |||||||:::
Db    466 TCTTTTTCTGAAGAA--------------------------------------  480

Qy    177 GlyLeuValAlaAsnMetIleAlaGlyLysAsnProPheLysMetProGlnMetArg  196
```

Db   1318 -CCAGCAGGAGAGTGTCAGATGAGAGAGTGTCTCCTGCTCAGAAAGCTCACTGCC 1376
                |||| ::: |||
Qy   481 --------------------------------AAAGGAGAATCTGATGATGAAGCCG-------AGG 510

Qy   525 oThrAlaLysAlaGluMetIleAspAlaLysValPheGlnAspIleGluArgProIl 545
Qy   197 LysAlaGlnAlaAlaProSerValPheGlnIleAlaLeuAlaLeuAlaLysArgAlaMetLeu 216

Db   1377 TGAATGCTCCACCAG--------------------AAGGGTGTTGAAAGTGAGCGTGAAAAGTCTGC 1424
Db   511 AAAGGGGAAAACGATCATCCAGAGTT---AGACAGCAAAATCTAAACTCCCTGAGTAC 567

Qy   545 ePro---ProLeu--------PhePheGlu-ProLysGlyArgHisThrArgLeu- 559
Qy   217 GlyLysAsnAlaProValAlaGlyValGlyLysGlyGluGlnArgMetMetAsnArg 236

Db   1425 TCCCCTCCCTTAACAGTGAGGAGGTTCGCACCTGCCAAAGGCATCACTGAGGAACCTAT 1484
Db   568 AGTCAGACTGCTGAGGAGGAAGAGAGGATCAAGAGACACCCATCTAGAAACCTGAGGGTCAGA 627

Qy   560 --------ArgTrpThrGlyAlaAsnGluLysGluIleProGlyLeuGlySerArgP 576
Qy   237 ValAspGlnArgMet--------GlnGlnArgGlnGluGlnGluAspGluAspAsp 253

Db   1485 GAAGAAACAGTCTTTGGAACAGAAGGAGGGCGAGAGCCTCCCCATGCTCTCTTCCCAGA 1544
Db   628 GCAGATCGAAATTAAAAATAGGAGGAAGAAGAAGAAGAAGATGAC 687

Qy   576 heIleLeuProSerLeuAspThrMetPro 586
Qy   254 AspAspLeuProGluAspGluValProArgArgSerSerAspGlyGluProGlnSer 273

Db   1545 GCACAGTGGGAAGCAGTCAGCCGACTCATCCT 1576
Db   688 GATGATGAAGGAGGAGGAGTA----------------GAT 714

RESULT 15
BW470454/c
LOCUS   BW470454   837 bp   mRNA   linear   EST 11-JUN-2004
DEFINITION   BW470454 Nori Satoh unpublished cDNA library, juvenile whole animal
              Ciona intestinalis cDNA clone cijv416e07 3', mRNA sequence.
ACCESSION   BW470454
VERSION   BW470454.1 GI:48606318
KEYWORDS   EST.
SOURCE   Ciona intestinalis
  ORGANISM   Ciona intestinalis
              Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea; Enterogona;
              Phlebobranchia; Cionidae; Ciona.
REFERENCE   1 (bases 1 to 837)
  AUTHORS   Satou,Y., Nakayama,A., Shin-i,T., Kohara,Y., and Satoh,N.
  TITLE   Expressed genes in Ciona intestinalis (2004b)
  JOURNAL   Unpublished (2004)
COMMENT   Contact: Nori Satoh
              Department of Zoology
              Kyoto University
              Sakyo-ku, Kyoto, Kyoto 606-8502, Japan
              Tel: 81-75-753-4081
              Fax: 81-75-705-1113
              Email: satoh@ascidian.zool.kyoto-u.ac.jp.
FEATURES   Location/Qualifiers
     source   1..837
                     /organism="Ciona intestinalis"
                     /mol_type="mRNA"
                     /db_xref="taxon:7719"
                     /clone="cijv416e07"
                     /tissue_type="whole animal"
                     /dev_stage="juvenile"
                     /clone_lib="Nori Satoh unpublished cDNA library, juvenile
                     whole animal"
ORIGIN

Alignment Scores:
Pred. No.:   0.00045   Length: 837
Score: 150.00   Matches: 65
Percent Similarity: 40.89%   Conservative: 45
Best Local Similarity: 24.16%   Mismatches: 87
Query Match: 4.56%   Indels: 73
DB: 5   Gaps: 11

US-10-736-868-2 (1-643) x BW470454 (1-837)

Qy   145 AspAlaLeuArgThrGlnArgAsp--------MetGluAsnThrAspProSerGlu 160
Db   837 GATGCCCTGCGCAGCTGTCTAAGAACGCCTGTCTCCTTTGTCAGTTGTCAGTACT- 778

Qy   161 GlnIleAlaLysAlaValMetAspLysPheGlnThrGlnIleLeuThrArgAspGl 180
Db   777 GAAGTCGGCAGTGATGCCAACCAAAGTTGCCGAGCCAAAAATTGCCAGT 718

Qy   181 Asn--------MetIleAlaGlyLysAsnProPheLysMetProGlnMetArg 196

```
Db  717 GATGCCAAACCAAAAGTTGTCAGTGATGCCAAACCAGTGGCGGCCAGTCGATGCCGAAACA 658

Qy  197 LysAlaGlnAlaAlaProSerSerValPheGlnGlnAlaLeuAlaLeuGlnAlaMetLeu 216
             |||||||||            :::||||||| |||  :::
Db  657 AAG---GACGTGGCCACCATCAGACACGAAACGGCTGCATCGAAGGTG------ 607

Qy  217 GlyLysAsnAlaProValAlaGlyArgGlyArgGlyGluGlu------------------ 229
             |||  |||         :::  :::
Db  606 ------GAAGCGGATGTGACGTCACAGAGGAAGACCACCACGATGTTTAATGACGAG 553

Qy  230 ------GlnArgMetMetAsnArgValAspGlnArgMetGlnArgMetGlnGluLeu 246
                 :::||||||||:::  :::|||||||||||| |||
Db  552 GAAATGTTCGAGAGAATGATGAAAGACAGAATGATGAAACGCATGCAACAAATCGAGGAC 493

Qy  247 GlnGluGluAspGluAspAspAspLeuGluAspGluAspValProArgArgArgSer 266
             :::::|||||||||||:::::::::    |||||:::::
Db  492 GAAGATGAAGACGAAGACGAAGAAGAA---GAAGATGAGAATACG--------AGAGAC 445

Qy  267 SerAspGlyGlyProGlnSerGluAlaGlyHisGlnArgArgAsp-LeuAlaArgArgLe 286
             |||     :::     |||  |||    :::|||:::
Db  444 CATGATCAGGATGACCATGTTGAGGAGGAGGCCACAGAGGATGAAGAGGAT 385

Qy  286 uLysSerSerProArgLeuLysGluLeuLeuGlnAsnAlaGluValGlnSerLeuLeuSe 306
             :::        |||  |||        ||
Db  384 GAGGAAGAAGAAACGAAGAAAGATGAGACAACAGAGA----------------------- 348

Qy  306 rTyrGlnArgMetArgAspSerProLeuSerLysArgArgProLeuLeuAlaMetAsnAspGl 326
             :::|||                  |||||||||:::::::|||
Db  347 ----AGAAGGAGAATGACGTCACCGCTGTGAAACCAAAAGCCG--------- 309

Qy  326 uAspGluSerAlaPheArgAlaMetGluAlaArgAlaLysLeuAspGlnLysSerGlnLe 346
                          :::||||:::::||||||||||||||
Db  308 ----------------ACGCCCAGGTTGAACCAAAGGTCACAGCT 280

Qy  346 uValLeuGlyLeuHisGlyPheGlyGluSerAspAspAspGluGlnGluAspGluAsp 366
             |:::             |||||||||||||||   :::
Db  279 GATAAAGCAAAAACAC----------CAAAAGTC-TCCAACGAAGACGAGATGGTGAAGAA 230

Qy  366 nLeuIleAspProSerGluAsnSerPheArgArgAlaProLeuArgLeuSerSerGlyPh 386
             |||  |||
Db  229 ACTTCGAGAT------------------------------------------CT 218

Qy  386 eValGluLysLeuLysSerAsnAsp 394
             ::::::::::::::||||||:::
Db  217 TCTCCAACGTATGAGAAGCAACGAG 193
```

Search completed: July 18, 2005, 16:09:04
Job time : 5650 secs

THIS PAGE BLANK (USPTO)